DOCUMENT RESUME

ED 044 275 SE 009 335

AUTHOR	Scott, M. Gladys, Ed.
TITLE	Research Methods in Health, Physical Education, Recreation.
INSTITUTION	American Association for Health, Physical Education, and Recreation, Washington, D.C.
PUB DATE	59
NOTE	544p.; Second edition
AVAILABLE FROM	American Asso. for Health, Physical Education, and Recreation, 1201 16th St., N.W., Washington, D.C. 20036 ($6.00)
EDRS PRICE	EDRS Price MF-$2.25 HC Not Available from EDRS.
DESCRIPTORS	Curriculum, *Health Education, *Physical Education, *Recreation, *Research Design, *Research Methodology, Sampling, Statistical Analysis

ABSTRACT

Methods for designing and conducting research related to health education, physical education and reaction are described. The use of libraries, the selection and definition of research problems, principles of statistically sampling, and data collection devices are presented. Instructions for the construction and validation of tests as well as techniques of data analysis are included. Descriptive methods, action research, and historical and philosophical methods, and experimental designs are described. Techniques for writing research reports, including advice on style and presentation of data, are given. Interaction of research and the development of the curriculum are discussed. (AL)

RESEARCH METHODS

in health physical education recreation

2ND EDITION

Edited by
M. GLADYS SCOTT
for the
RESEARCH COUNCIL

AMERICAN ASSOCIATION FOR HEALTH, PHYSICAL EDUCATION, AND RECREATION | WASHINGTON, D. C.
1959

Library of Congress Catalog Card Number 59-15016

First Edition—1949
First Edition, Revised—1952
Second Printing—1956
Second Edition—1959
Second Printing—1961
Third Printing—1964
Fourth Printing—1965
Fifth Printing—1966
Sixth Printing—1967

Price: $6.00

Discounts on quantity orders: 2-9 copies, 10%; 10 or more copies, 20%

PREFACE

EXISTING RESEARCH TECHNIQUES ARE BEING IMPROVED CONSTANTLY and new research techniques are being developed. The beginning graduate student in health, physical education, and recreation should find this book a helpful introduction to present-day methods and techniques of research. It is hoped that the students who seriously desire suggestions for solutions to problems will find help in this edition.

The emphasis in the writing of the chapters has been placed upon the application of the research methods and techniques. The complete procedures for methods and techniques will not be found in all instances, but enough guidance will be found for a student to obtain a basis for further development of solutions to problems The research methods and techniques described are used in our fields of health, physical education, and recreation, as well as in the fields of education, anatomy, audio-visual aids, physiology, and psychology. Help beyond that given in this book may be found in research done in our fields and in other fields, as well as in other research methods books.

This Second Edition is a completely new version of *Research Methods Applied to Health, Physical Education, and Recreation*, which was originally published in 1949 as a project of the Research Section of the American Association for Health, Physical Education, and Recreation. Most of the chapters have been completely rewritten. A few chapters have been revised and brought up-to-date. Appreciation for the use of any of the original material is expressed to the contributors to the 1949 edition who were: Dorothy Ainsworth, Kenneth D. Benne, Carolyn W. Bookwalter, Karl W. Bookwalter, David K. Brace, Freeman Brown,

H. Harrison Clarke, Louise S. Cobb, Thomas K. Cureton, E. C. Davis, Robert L. Ebel, Anna Espenschade, Nicholas Fattu, Esther French, Ruth Glassow, Franklin Henry, Jack E. Hewitt, Pauline Hodgson, Alfred W. Hubbard, Laura Huelster, Rheem F. Jarrett, Hardin B. Jones, Peter Karpovich, Louis Keller, Ellen Davis Kelly, Hyman Krakower, Leonard A. Larson, C. H. McCloy, Howard V. Meredith, Walter S. Monroe, Marjorie Phillips, Margaret S. Poley, Elizabeth Rodgers, Helen R. Russell, M. Gladys Scott, H. L. Smith, Ralph B. Spence, Arthur H Steinhaus, R. H. Stetson, and Jesse F. Williams.

This Second Edition has been a project of the Research Council of the American Association for Health, Physical Education, and Recreation. The responsibility for each chapter rests with its author(s). Copies of the preliminary manuscript were circulated to the eight members of the Steering Committee, who read the material and offered constructive criticisms. The administrative responsibility for handling the manuscript in its various stages was delegated to Dr. M. Gladys Scott, who served as chairman of the Steering Committee.

Acknowledgment is made to Dr. Elena Sliepcevich for her services as consultant on health education materials, to Dr. Chester W. Harris for his services as consultant on the chapter on Populations and Samples, and to Dr. Joseph Brozek for his services as consultant on the section on Applied Physiology. Recognition is also given to Dr. Carl A. Troester and his staff and to the Board of Directors of the American Association for Health, Physical Education, and Recreation for the support of the project.

THE STEERING COMMITTEE

M. Gladys Scott, *Chairman*
Carolyn W. Bookwalter
Karl W. Bookwalter
H. Harrison Clarke
Laurence E. Morehouse
Marjorie Phillips
Arthur H Steinhaus
Raymond A. Weiss

CONTENTS

1. WHY THIS RESEARCH? 1
Arthur H Steinhaus

2. LIBRARY TECHNIQUES 20
Carolyn W. Bookwalter and Karl W. Bookwalter

3. SELECTION AND DEFINITION OF A PROBLEM 39
Karl W. Bookwalter

4. POPULATIONS AND SAMPLES 71
Frances Z. Cumbee and Chester W. Harris, Consultant

5. TOOLS FOR OBTAINING DATA 99
   - Typical Research Tools in Education 100
     Esther French
   - Attitude Scales 114
     Charles C. Cowell
   - Sociometric Techniques 121
     Charles C. Cowell
   - Photography 128
     Alfred W. Hubbard

6. LABORATORY PLANNING AND INSTRUMENTATION 148
Frank Sills

7. TOOLS FOR ANALYZING AND PRESENTING DATA 170
H. Harrison Clarke and Marjorie Phillips

8. CONSTRUCTION OF TESTS 213
Raymond A. Weiss and M. Gladys Scott
   - Written Tests 215
   - Physical Performance Tests 237

9. DESCRIPTIVE METHODS 252
   - The Survey 252
     Elwood Craig Davis
   - The Case Study 264
     Lawrence Rarick
   - The Genetic Method 269
     Lawrence Rarick

10. THE EXPERIMENTAL METHOD 277
Marjorie Phillips and Leonard Larson

11. LABORATORY RESEARCH 300
Anthropometry 300
Margaret Fox
Body Mechanics 315
Margaret Fox
Kinesiology and Activity Analysis 326
C. Etta Walters, John Cooper, and Olive Young
Methods of Instruction 350
M. Gladys Scott
Applied Physiology 360
Marjorie U. Wilson and Josef Brozek, Consultant
Applied Psychology 372
M. Gladys Scott

12. RESEARCH AND THE CURRICULUM 397
Research and the Curriculum in Physical Education 398
Louis E. Alley
Research and the Recreation Program 423
Ralph H. Johnson
Research and the Curriculum in Health Education 437
Wallace Ann Wesley

13. ACTION RESEARCH 453
Camille Brown and Ruth Abernathy

14. THE HISTORICAL METHOD 465
D. B. Van Dalen

15. PHILOSOPHICAL METHODS 482
Eleanor Metheny

16. WRITING THE RESEARCH REPORT 502
Carolyn Bookwalter, Karl Bookwalter, and Marjorie Phillips

INDEX 532

ACKNOWLEDGMENTS

The chapters in this book were written by members of the Research Council of the American Association for Health, Physical Education, and Recreation, who are outstanding research persons in the field of health, physical education, and recreation.

The Research Council greatly acknowledges the contributions of the following authors:

Ruth Abernathy, University of California at Los Angeles; *Louis E. Alley*, State University of Iowa; *Carolyn W. Bookwalter* and *Karl K. Bookwalter*, Indiana University; *Camille Brown*, University of California at Los Angeles; *H. Harrison Clarke*, University of Oregon; *John Cooper*, University of Southern California; *Charles C. Cowell*, Purdue University; *Frances Z. Cumbee*, University of Wisconsin; *Elwood Craig Davis*, University of Southern California; *Margaret G. Fox*, State University of Iowa; *Esther French*, University of Michigan; *Alfred W. Hubbard* and *Ralph H. Johnson*, University of Illinois; *Leonard A. Larson*, New York University; *Eleanor Metheny*, University of Southern California; *Marjorie Phillips*, Indiana University; *Lawrence Rarick*, University of Wisconsin; *M. Gladys Scott* and *Frank Sills*, State University of Iowa; *Arthur H. Steinhaus*, George Williams College; *D. B. Van Dalen*, University of Pittsburgh; *C. Etta Walters*, Florida State University; *Raymond A. Weiss*, New York University; *Wallace Ann Wesley*, American Medical Association; *Marjorie U. Wilson*, University of Minnesota; *Olive G. Young*, University of Illinois.

CHAPTER 1

Why This Research?

ARTHUR H STEINHAUS

THE EMPLOYMENT OF EDUCATIONAL METHODS TO ACHIEVE BETTER health, greater fitness, and the fuller enjoyment of life is an art, not a science. When this is done in the interest of humanity with reasonable likelihood of meeting certain needs of mankind it constitutes the practice of a profession. Those practices which we have come to include under the headings of health education, physical education, and recreation constitute an important specialty or branch of education, the oldest of the professions.

OUR PLACE

In this area where education becomes interested in health it borders most closely on another great profession, namely, the art of the practice of medicine. In fact, on this border there sometimes has been understandable but needless confusion. It is understandable because both professions have the same goal—the improvement of man's physical, mental, and social well-being; and both professions draw their factual information from the same pools of knowledge—the basic physical, biological, and social sciences. It is, however, a needless confusion because these professions should not differ in their ultimate goals nor in their sources of knowledge, but should differ only in the practices they employ to bring these knowledges into the service of mankind.

We in education employ the methods of education. Our colleagues in the medical services employ the methods of medicine. At the risk of oversimplification may it be suggested that medicine in common with other services to mankind, such as public health and engineering, does things for people; whereas education, strictly speaking, does nothing for people, but instead helps people to do things for themselves. Consequently, the methods of education are as different from those of medicine as are the practices of farming or tree culture different from those of bread making or carpentering.

Research in our fields may appropriately interest itself, (a) in basic research which aims to enlarge the pools of knowledge common to all professions, or (b) in applied research which aims to discover the best ways of using this knowledge in the practice of our professional art. Sometimes the distinction between basic and applied research is not as simple as this sounds. It is safe to assume that most of us will do applied research.

THERMOMETERS AND THERMOSTATS

The thermometer records temperature; the thermostat does something about temperature. Every thermostat needs a good temperature recording device. It needs also many other accurately designed, carefully fitted parts and, in addition, some source of power to move valves. When this power is properly released and controlled in accord with predefined objectives of desirable temperature, the thermostat performs its function.

The sciences and scientists, like thermometers, observe, measure, and record. The professions employ the findings of science to do something for mankind. Always, the professions must draw on many sciences for the best available facts lest they fail mankind. In 1927 the great St. Francis Dam near Los Angeles gave way. The commission that probed the cause of this catastrophe reported that the perfectly constructed, modern concrete structure had been built on a rock bed of mica schist in an area of geologic faults. Any geologist could have predicted the disaster. Because the science of geology had not been permitted to contribute its pertinent facts, the engineering profession's efforts at St. Francis were a disservice to mankind. The commission strongly urged the inclusion of geology in the curriculums of engineering schools.

The profession of education and its specialty, physical education, may be of far greater significance to man than is engineering; but few people realize this—which is fortunate for us. What would become of us if people attributed to education all the great disasters for which it is responsible? The list would include depressions, congressional filibusters, an assortment of world wars, and diplomatic failures at Lake Success, as well as their backyard counterparts: ignorance, delinquency, prejudice, ill health, and foul play. What do we lack?

If this seems but a pompous beginning for a manual on research, let its intention be clear. The greatest danger that besets the professional worker who engages in research is that while absorbed in examining details, he may lose his professional perspective; and even worse, his orientation that gives purpose and direction to all his endeavors.

This is not to deprecate the complete mastery of detail. It is rather to give purpose to detail. The easier airplane view will never master the problems of the forest. Too few of us are willing to toil the hard hours in the depths of detail which alone will produce solid foundation for a strong profession. Such constitute the sad array of lazy, unprepared workers. A more pathetic figure, however, is the nearsighted fusser with statistics who substitutes mere toil for directed effort and is so engrossed in each figure that he does not read the score.

The purpose of this manual is to encourage and assist intelligent research directed toward worthy goals. Such activity will at once advance the profession and the professional worker.

In its early beginnings a profession must draw its personnel from related professions and its principles from related fields. In time it will generate more and more of its own personnel and guiding principles. This desirable evidence of growing maturity is, however, not without danger. Confined to its own sources the profession may stagnate. Always some of our best graduate students should find their ways into the graduate divisions of the natural and social sciences of our great universities, there to search the latest findings and the crispest methods of pure research in the older disciplines. They should go well prepared and alert to "pick the minds" of the master workers in these fields in order to bring back to physical education that which is new, stimulating, and helpful. Such students should find challenges in what this manual leaves unsaid as well as in the bare spots of our knowledge that it points out. For techniques of research they will look elsewhere. For the larger number of our graduate students, whose interest is rightly in applied research as a way of becoming more intelligent and productive professional workers, this manual is "tailor made." For those who face a research task as the final obstacle before that higher degree and the promotion that depends on it, this manual is a "godsend." It might direct even such unwilling steps to useful ends.

But let nothing be said to cause this latter group to take on greater feelings of inferiority. Research is important; it is imperative; but it is not all! The majority of our most inspiring teachers and most efficient administrators are miserable researchers. This is true also in medicine and the other professions. If all of us were thermometers, who would operate the valves? Even more crucial, who would say, "It's too hot here, let us reset the thermostat"? Finally, who would dare stick out his neck to say, "Let us create a new thermostat"? Obviously such functions are as essential for the creation and operation of useful programs for a city system as they are for the attainment of a suitable room temperature.

The best thermostat, speaking for our profession, is one in which each part is conversant with the ways of working and purposes of the other, and is appropriately influenced by the special contribution of the other. An illustration will serve to clarify this point. All of us at sometime have had to learn the basic laws of health and many of the games and skills that constitute sports and recreation activities. We have also gained some understanding of children and adults. Some of us have become expert performers, some expert teachers. Others of us must supervise or administer a diverse range of programs and rarely operate at the performing or teaching level. Nevertheless, we are better workers because we know that which must be taught and those who must be taught, and we understand the tasks and the tribulations of the teacher. Still others of us may come into a situation where we must completely revamp a program or an entire system.

DISCIPLE, PRACTICAL WORKER, STUDENT

In this illustration one thing is missing. How do we know that we are teaching the right things in the right way to the right people at the right time? In fact, how do we know anything is right or best? People answer these questions in different ways. Some say it is right because they themselves were taught that way. Such people who tacitly accept the authority of a leader are *disciples*. Others say it is right because it works. These are the *practical workers*. Maybe these persons are right, but without some measurement of their results they have only their own intuition and perhaps the approval of others to guide them. Often they tend to be dogmatic. For much of our practice this may be as far as we can go at present.

Still others seek the largest base of experience by which to test all practice. They discover what others have found and practiced; they devise ways of examining the results of their practice by methods that will exclude their own prejudices; they find ways to study themselves and others as objectively as possible; they learn to judge the degree of accuracy of their findings to know how seriously to accept them; and when they must make decisions that reach out beyond the facts they have been able to establish, they employ reasoning disciplined by all the available facts. Even then, they do not take their projected judgments or hypotheses too seriously. They know the difference between a hunch or guess; a good working hypothesis; a well-supported theory, law, or doctrine; and an established fact. Such people are *students* of their profession; their practices are scientifically based; they are imbued with the scientific spirit. They are all this whether their day's work finds them leading play in the nursery, coaching a football team, classifying pupils for intramural competition, making an administrative decision, sitting on a community council, or baking a cake. When such "on the job" practical searching and experimentation is subjected to disciplined record keeping and rigorous control of procedure, perhaps in a sample or pilot-type situation under actual working conditions, it is often called clinical or *action research*. Needless to say, such activity demands a high level of research leadership lest it degenerate into a mere proving of prejudices under the exercise of politically held powers.

Any person who accepts the public trust implicit in the position of a professional worker is morally bound to true up his practices with the findings of our best researches and his methods of working with the recognized habits of modern science. These he can learn best by intelligently exploring the researches of others and consciously practicing their methods.

THE METHODS OF RESEARCH

The methods of research, though endlessly different in their specific application, are fundamentally simple. Essentially they comprise four steps: observation; recording, organizing, and treating the observed data; generalization to the formulation of a theory; and testing the new formulation with further observations.

Historical Research. Many things have happened before the researcher comes on the scene. In such instances he must depend on observations made by others who lived before him, and variously recorded in personal files, letters, minutes of meetings, and other contemporary documents of all kinds—even in the memories of friends or relatives. This is *historical research.* If it centers on the life of an individual it is sometimes called *biographic research.* It is obviously limited by the availability of materials, and begins with their discovery. The historian has refined the methods of discovery and validating the authenticity of his raw data, but essentially his activity is confined to the second step in the above list, i.e., the recording, organizing, and treating of his data, which are the observations made by others. Often such research covering the development of concepts over a long period of time provides clues to the formulation of new generalizations, which lend themselves to further testing. This bringing together and reworking of older ideas and findings is itself a form of research sometimes called *collation* or *integration.* It is also called *library research* and should in fact be an early step in every research program.

Observational Research. Many phenomena such as the movement of stars, the weather, and the behavior of our fellow human beings are occurring contemporaneously, often in our very midst, subject to regular and even continuous observations. Often these phenomena are entirely outside our control. We must take them as they come and sometimes it is a long wait. Thus, the sun's corona has been under observation by astronomers for less than two hours all in all. The late Professor David Todd of Amherst probably witnessed more solar eclipses than any other scientist, yet his number totalled only eleven and during seven of these, clouds obstructed his view. Phenomena such as the weather, the growth of children, and the behavior of nations, though much more commonly observed, are nevertheless almost as completely outside our control. We may count, weigh, measure, average, and chart them. We may submit the findings to endless mathematical treatment but always we must wait for the phenomena to happen.

This is known as *observational research.* It employs questionnaires to gather factual data, or, "opinionnaires" as some have called them, to gather opinions, often by mail and from a great number of persons or institutions. This would be called a *broad survey.* At other times, such data are gathered by visitations and interviews employing interview schedules, checklists, testing, measurements, or other more intensive case study procedures. It is then called *analytical survey.* Sometimes the data so observed and collected are used in the development of scales for rating and scoring achievement and for comparing individuals or groups. Such normative procedures may be called *normative surveys.* At times data are subjected to other statistical procedures to determine the extent to which two or more kinds of observations made on an individual or a group have a tendency to be related or found together (*correlation*); to determine the extent to which one or several observed items may be causally related to another and therefore usable to predict the other (*causal analysis*); and finally to determine the nature, number, and relative importance of several causes known or unknown, that together produce one result (*factor analysis*).

The validness of conclusions drawn from any series of observations depends on the representativeness of the raw data. Sometimes it is possible to place all members of a group under observation. This is the nature of any *census* study. More commonly, for practical reasons, this is impossible. Then it becomes important to get a truly representative *sampling.* The determination of such a sample population for study is itself a critical procedure for which special methods are available. There are also statistical procedures which give an indication of the degree of certainty that the results derived from computations based on one sample are likely to be identical with results gained from similar observations on another sample (*probable error*).

Thus statistics in essence provides a mathematical control over an uncontrollable experiment being carried out by nature. This ingenious use of mathematics is the quintessence of abstraction from concrete observations. Just as whipped cream, no matter how bulky its billows of foam, is never better than the cream of which it is made; so the abstractions of statistics, no matter how impressive they seem, are never better than the initial raw data that went into the formulas. Many other pitfalls beset the user of statistics, so that often such departures from raw data—the terra firma of accurate observations—sink in the quicksands of bemuddled thinking, far below the level of sound abstraction.

Experimental Research. Thus far we have considered research in which man's efforts are limited to observing phenomena when and if they occur about him, and taking them as they come. Fortunately some phenomena can be made to happen at the will of the observer. Thus the chemist, by mixing the proper ingredients at the right temperature, etc., in his test tube, can reproduce endlessly and at will what nature may do only rarely and in remote corners. Similarly, by completely controlling the diet and other environmental factors of an animal, the physiologist is able to study at will the effect of a single vitamin or amino acid on growth. By observing thousands of flies in scores of generations of predetermined and forced matings, the geneticist establishes the laws that govern the mixing and resorting of the genes that determine what we inherit. This management of the several components which together cause or determine a phenomenon under observation is the mark of the *controlled experiment* in *experimental research.* It is the latest accomplishment of science. Though best developed in physics and chemistry, it has also been profitably employed in the biological field beginning with Harvey in 1628.

Not all phenomena are subject to study by all three of these methods, i.e., the historical, observational, and experimental. Nor is this necessary in order to have an accurate science. For example, astronomy, one of the most accurate sciences, will never become an experimental science in the complete sense of the term as here used until solar eclipses can be produced at will. Psychology has only recently entered the experimental field. It is most doubtful if the social sciences will ever become truly experimental in the sense of the controlled experiment. This is no reflection on social scientists; it is a comment on the extreme complexity and uncontrollability of all but the simplest social phenomena.

PREDICTABILITY THE REAL MEASURE

The real measure of a science is its ability to predict. Predictability is based on the assumption that ours is a world of "law and order." This is the fundamental *belief* of scientists and is inspired by a *faith* in the inescapable relation between cause and effect. These may be called doctrines, or better, the charter or *sine qua non* of science. Without them there would be no science, and research would be an organized wild goose chase.

This needs illustration. Aristotle said, "Jupiter rains not to grow corn but of necessity." Even Jupiter cannot help it. Once all the preliminary causes and sequences have transpired he must rain. Today we explain the cause of rain in terms of sudden cooling of warm moist air. Cooling is due to wind currents moving cold air into warm air or warm air into a cold spot. Wind is caused by the unequal expansion of gases under the influence of sun, etc. When the proper combination of conditions prevails the water vapor condenses (perhaps around tiny particles of dust whose presence is also accountable) and it rains. It can't help itself. Man has learned to observe many of these preliminary steps. His hundreds of meteorological stations make and record thousands of observations. His high and low pressure maps chart these findings into a composite picture of the atmosphere's behavior. If man is sensitive to a sufficiently complete number of factors and if he has the skill, acquired from previous experience, to interpret their interactions accurately, he predicts the weather.

In astronomy, where measurements have reached greater perfection, man is able with split-second precision to predict solar eclipses, the time of sunrise, and the time schedule of the evening star. Man has not yet learned to change the weather very much and probably never will be able to hurry an eclipse. But if he were able to change them, it would be man the *engineer*, not man the *scientist*, who did it. The final test of science is not its ability to change phenomena, but its ability to predict them—not to control the universe, but to know it. Such knowledge may then be used by man to serve his ends. If he cannot change this world, he may change himself like Mark Twain's *Yankee in King Arthur's Court* who saves his life by awing angry savages into worshiping him when he "invokes" a solar eclipse at the psychologic moment which he himself created, after consulting his almanac and wrist watch.

Ofttimes we are convinced of predictability even though we are unable to predict. In this sense, predictability expresses a faith which authorizes the expenditure of research energies to uncover relationships that may guide professional practices short of complete predictability. Imagine, for example, a raindrop several hundred feet above the earth's surface. Can you predict the spot on which it will land—let us say to within an inch? You may say, "No, and no one cares." But let us see. Do you agree that its course is determined by gravity and many other more variable factors, and that if you knew the strength of every gust of wind, its exact direction and its duration, if you knew how the drop was deformed and therefore the resistance it would offer to the next gust from another direction, and so on and on, will you agree that if you could take the time and had the facilities for measuring all this, then you could predict where it would land? If so, then you agree that the raindrop's course is *predictable* and *researchable* whether or not anyone will ever predict exactly where any one drop will land.

Concerning the effects of sports on the body of man, we have at present sufficient data and fairly defined principles to permit some prediction. Concerning the effects of sports on the mind and spirit of man, we have beliefs and convictions but few facts. We are here in the predicament of the raindrop analogy. There are effects, and there is rhyme and reason about them which justifies the assumption of predictability. Our difficulty stems from the fact that we do not as yet have sufficiently adequate devices for observing and measuring changes in all areas; and where we can measure we are confronted by a terrifically complex array of possible causes. To illustrate: It is easy to determine whether a season of football or weight lifting under specified conditions has resulted in increased body weight or caused muscles to grow. It is possible to predict that the increase in muscle size will be roughly proportional to the intensity of work, i.e., the amount of work done in a unit of time. Further, we know that this will be true, though in varying degree, depending on body type, of all healthy persons so engaged. This begins to sound scientific. In contrast, who is able to predict just what interracial attitudes, what standards of honesty, or what appreciation of infinite values will develop in consequence of a season of sport? Even though most of us are convinced that such changes do occur, who knows their direct causes? Is it the smile of a competitor, the odor of his perspiration, or his dogged perseverance that turns the trick? And does it turn the trick in the same direction for each participant—or for each spectator? Obviously, in these areas we cannot predict even though we are convinced of the existence of the kind of law and order that permits the assumption of predictability. It is our inability to comprehend and measure all of the multitudinous variants which here act and interact, in different combinations for each person, that has to date delayed the research necessary to gain scientific mastery.

Also in these areas it is not man the scientist who will bring about change in body size, contour, skill, or personality of people; but man the teacher, coach, or counselor who, using the best available facts from research and tested experience, will work with people to produce changes. The skills, interests, temperament, and even motivations demanded of the discoverer of facts and those demanded of the user of facts are far from alike. This probably explains why the best research scientist may not be a good teacher; and conversely, why many a topnotch teacher is poor at research. A sincere respect for truth and the highest level of integrity, however, must be qualities common to both.

FACTS AND THEORIES

Research is the scientific method for finding answers to questions. When we are insufficiently informed we often find it convenient to formulate tentative answers based on the available facts. Such tentative answers or generalizations are called hunches, guesses, hypotheses, theories, laws, or principles depending on how sure we are of them. They are important tools of thought because they help to place the known facts in proper relation to each other, show up the limitations of available facts, and most important they provide something to test or shoot at, the doing of which may uncover further facts, and thus hasten the finding of the true answer to the question.

A homely illustration will serve to clarify the interrelation of facts and theory and the special importance of each. Let us say that for some reason it is imperative that you know the whereabouts of Mr. Brown. Your question is: Where is Mr. Brown? From the observations and recordings made by others before you, you are in possession of Mr. Brown's street address and a city guide containing a city map, a register of streets, and street car directions. Library research in the telephone directory discloses that Mr. Brown has no listed telephone; consequently, you use the available records to find your way to his home. You knock on the front door which you observe stands ajar. There is no reply. You push the door open and look around the hall. Though the night is cold, you see Mr. Brown's coat on the hall tree—the hat is missing. Through another door you see the dining room table covered with dishes. Closer examination discloses warm coffee in the cups and plates but half emptied. The chairs are pushed back, one is lying on its side. These are all facts. You say to yourself, "Mr. Brown left the house in a hurry." That is your first generalization or guess concerning the whereabouts of Mr. Brown. You look around and your glance falls on another door half open leading into a bedroom. On the bed, you see a woman apparently unconscious. She groans. Now you theorize: "Aha, Mr. Brown went to the theater"—do you? Perhaps you will want to assure yourself that the unconscious woman is Mrs. Brown before you decide—but very likely you will conclude that taking all the facts so far gathered into account, the best guess concerning the whereabouts of Mr. Brown is that he has hurried out to get a doctor. That is your theory. But it is only a theory, no matter how plausible it sounds; and it must remain a theory until it is verified. True, it may be a fact. *The point is you don't know whether or not it is a fact.*

There are now two courses open to you in your search for Mr. Brown. You may decide to sit down and wait for his return; or you may adopt your guess as a working theory and start out for the nearest doctor's office. You are energetic and a little impatient and therefore start out in search of the nearest drug store. Here you inquire concerning doctors' offices, Mr. Brown, Mr. Brown's doctor, etc. You are told that the doctor upstairs is on a hunting trip so you start for the one who is said to live a block down the street. Before reaching the next corner you run into Mr. Brown who is carrying a prescription slip and in a moment of conversation your theory is confirmed. Further, your theory has helped you find Mr. Brown several minutes before you would have found him had you waited at his home. But let us say that you had not met Mr. Brown but instead you had run into a crowd of people just dispersing and the police patrol wagon starting away from the scene. The crowd is talking about a drunk Mr. Brown who had got into a fight. Their description of Mr. Brown tallies with that of your friend. Here are new facts. You cannot now go on unquestioningly holding your theory that Mr. Brown has gone to fetch a doctor. In your confusion you bump into a mutual friend who definitely settles the identity of this Mr. Brown as the one you are looking for. Should you now continue to claim that Mr. Brown went to a doctor's office, the world would call you a fool and say that you lacked judgment. Someone might say you were unscientific.

Actually, you are failing to consider all of the available facts in the formulation of your theory. You have a closed mind. You are a standpatter refusing to adjust to more recently established facts. This is even worse than had you jumped to the conclusion that Mr. Brown had gone to the theater when you first entered his home! The theory as well as the facts has its place. Without a theory one merely sits and waits. The skill exhibited in developing a theory is called judgment. *We must not minimize the importance of the theory.* On the other hand we must not be fooled into confusing a theory with facts. Someone has said the facts represent the maps, charts, and logs of our journey. Theory is the compass which guides us in the seeking of further facts. Together they guide us over the high seas of human experience to our destination.

Now let us ask another question: How is muscular strength developed? Although it sounds almost as simple as "Where is Mr. Brown?" there is much more to it. Even after ruling out the diverting questions usually raised—"What do you mean by strength?" and "Why is strength necessary?"—more pertinent ones remain:

1. How can strength be measured in animals and man?
2. How is strength related to muscle size?

3. What happens in a muscle when it becomes larger and stronger?
4. What kinds of exercise will most rapidly develop muscular strength?
5. Is there a limit to the amount of strength that can be developed in a person?
6. If so, what factors determine the limit?
7. Are differences in the ability to develop strength related to constitutional type?
8. Are the commonly observed sexual differences in strength biologically or socially conditioned, or both?
9. Given two people with the same size of muscle why can one generate more strength than the other?
10. Why is a person often able to exert much more strength under hypnosis and in acute emergencies than ordinarily?
11. What series of tests will give the best picture of a person's over-all strength?
12. To what extent can a person's success in different sports or playing positions be predicted from measures of strength?
13. What strength criteria shall be used in the selection of military personnel for various responsibilities?
14. How much strength shall be required in the training of soldiers?
15. Is there correlation between physical strength and
    (a) resistance to disease?
    (b) mental health?
    (c) personality adjustment?
16. Is strength more readily developed in the young or in the mature organism?
17. Does strength developed in childhood persist throughout life?
18. How long does it last?

Anyone can think up many more questions and subquestions, each of which may be as large or larger than "Where is Mr. Brown?" In going through this list the reader may have found himself formulating answers to each question. Whether these answers were guesses, well-substantiated principles, or rank falsehoods based on superstition depends only in part on the preparation of the reader. Some of the questions even in this relatively simple field are completely unanswerable in today's state of knowledge, and the rest are answerable with greatly varying degrees of certainty.

Thus, questions 1 to 4 can be answered with reasonable certainty because of observations made under controlled, experimental conditions in physiological and histological laboratories. Answers for 2 and 4 are in addition anticipated by historical research and supported by clinical research.

Answers for 5 and 6 are broad generalizations based on histologic observations of the limits of hypertrophy, the diffusion rate of oxygen, and the inability of highly differentiated muscle fibers to multiply.

Answers for 7 and 8 are based on statistical treatment of clinical observations with some verification for 8 from controlled experimentation.

Answers for 9 and 10 are at the stage of working hypotheses or mere guesses based on observations made in psychologic laboratories and psychiatric clinics.

Answers for 11 to 14 are based on extensive observation under clinical conditions that have been treated statistically to determine which scores correlate with success. This is likely to culminate in a series of correlation indices, which are really theories or working hypotheses, of how important strength is to this or that. Unfortunately research too often stops here, without testing its theories. Only rarely does one find a study that tests its claimed ability to predict, for example, a man's time in the 440-yard run from certain measurements, by subsequently determining how closely other runners' actual times compare with the times predicted for them from various measurements and test scores. That the development of strength in the training camp will help ensure a soldier's success partakes of the nature of *predictability* in the same sense as does the path of a falling raindrop cited in an earlier analogy, and is obviously even more difficult to prove by actual prediction.

The answer for 15a is based on widely accepted generalities or principles from related professions, but inadequately tested by controlled experiment. Answers for 15b and 15c are, at best, good hunches based on isolated clinical observations.

Number 16 is answered only with moderate certainty from experimental observations on animals, insufficiently corroborated on man.

Numbers 17 and 18 can be answered with reasonable certainty from controlled studies on animals and from careful observations on man made in laboratories, gymnasiums, and hospitals.

From the hundreds of studies that contribute in one way or another to the answering of the above questions, three broad principles or laws have emerged. These may be designated the overload principle; hypertrophy of use and atrophy of disuse, or the principle of reversibility; and the principle of individual and sex differences. Such principles guide the formulation of a philosophy of physical education that is scientifically grounded when it advocates some segregated activities for older boys and girls, strenuous programs for the development of strength, the necessity of continuing an exercise program throughout life, negation of absolute standards of strength in favor of standards related to individual type and adapted to the everyday requirements of life, and a high degree of expectancy of prompt results from a carefully planned program of strength building.

REMAINING TASKS

The history of research in any field has some similarities to picking apples. The early pickers can get fruit without reaching very high. Because the lower branches are now bare some may think there is nothing left to pick. The facts are otherwise in our field. There remain more problems unsolved than any of us can imagine. True, some are on rather inaccessible branches. It may help to mention just a few.

At what stage in the development of an individual do muscle cells cease to multiply? Does exercise in any way modify the timing of this? Does the fact that all body proteins in man are renewed at least once in 160 days have any significance for training programs or training diets? Does a strenuous training program shorten this "turn over" period? Do any of the metabolites of exercise modify the colloid states of cell protoplasm in a way that might give basis for prolonging the general flexibility of youth? In what ways do exercise and the endocrine glands interact? Are there any observable changes in the composition of blood constituents in consequence of a period of recreational activity? Are there any objective measures of the amount of inner tension in a person that might be used to observe the progress of recovery from so-called "nervous states," and thus be used to justify recreational activities? What objective evidence, if any, can be found to throw light on the good or bad effects of highly emotionalized competition on the female organism? Is this in any way different from its effect on the male? What is the relationship, if any, between mechanical and chemical stresses exerted on muscles, cartilage, and ligaments in youth and the incidence of fibrositis, arthritis, and related conditions in later life? Is there any relationship between swimmers' cramps and the recency of alimentation, or blood sugar, or temperature, or state of fatigue? What effect does "drying out" to lose weight have on vital body functions. What habits of living such as diet, exercise, and emotional disturbance modify the clotting time of blood, modify other blood ingredients that may predispose to vascular diseases? What changes discernible by mental testing programs follow punishment inflicted to the head as in boxing? What effects of boxing may be revealed through intensive case studies of individual boxers? How do the vitamin and protein needs in strenuous training differ from those of usual living?

Even such a list as this is in no sense more than indicative of the vastness of the field for research. Obviously, some of the answers must first be sought in animal experiments. To the student who is thoroughly trained in biology and experimental psychology, experimentation on lower animals makes sense even though he does not assume that all findings are entirely applicable to man. The person who denies the validity of animal experimentation reveals a shallow understanding of the fundamental unity of living matter and the attendant implications for experimentation.

This list also points to the necessity for working more closely together with scholars in the fundamental sciences.

WHY THIS RESEARCH?

"Why This Research?" is answered in the analogy of woodchopping. Woodchopping produces both useful wood and a better woodchopper. Research must give to our fields the building materials of accurate facts and principles with which to construct sound practice and wise philosophy. It must supply ideas to kindle enthusiasm in our professional ranks and, in the public mind, a warm reception for our programs.

Research must also create for us a professional personnel that is expert in its attack on new problems, keenly alert to new opportunities, wisely guided in the efficient application of its energies, and disciplined with a fine humility that is fathered by confidence in one's power and mothered by an appreciation of one's limitations.

THE RESEARCHER HIMSELF

In the last analysis no work is better than the worker and the quality of research depends entirely on the knowledge, wisdom, and personal integrity of the investigator.

> If you are well informed of advances in your field, and yet endowed with a curiosity that breeds dissatisfaction with the present state of this knowledge,
>
> If you can ask significant questions and also formulate crucial methods for discovering honest answers,
>
> If you have imagination to conceive a dozen hunches, and at once the industry to explore each until disciplined wisdom points the one of choice,
>
> If you can concentrate on an issue, and yet be alert to happenings on the periphery,
>
> If you stick tenaciously to the rightness of your best hunch, yet possess the objectivity to treat it with detachment as though it were another's and stand ready to give it up when it becomes untenable,
>
> If enthusiasm and industry drive you to collect much data and you record with equal respect that which supports and that which negates your theory,
>
> If you are possessed of a fantastic memory for facts yet are willing to record them systematically as though you could not trust your memory,
>
> If your mind works with speed and accuracy, and yet you double check your calculations,
>
> If you are justly proud of your theory, yet humble enough to be led by facts,
>
> If you are "hell-bent" on proving your theory, and yet satisfied that disproving it is just as great a contribution to knowledge,
>
> If you have the courage to persist in the face of disagreement, and at once the patience to listen to the opposition,
>
> If you are endowed with energy for long hours of searching and have enough left to organize, tabulate, analyze, and publish your findings,
>
> If your mind is capable of holding the profoundest ideas and you have the understanding and restraint to express them in simple words even though you also know the big words,
>
> If you are eager to forge a reputation for yourself, and at once willing to acknowledge generously your indebtedness to the labors of others,

> If you are really capable of research, and your activity persists beyond your doctorate to the time when you must yourself supply both the time and motivation,
>
> If to all of the above you can give honest affirmation, you are better than most of your contemporaries and predecessors but you are none too good for service to health, physical education, and recreation.

* * * * * *

Let no reader of this book be unduly impressed. Though it is written by many of today's best minds in our fields, the writers will be the first to admit shortcomings. He would be a poor reader who could find no flaws in these pages. Such discovery will not discourage the authors, but will be welcomed by them as the mark of an intelligent rising generation. But if such a generation will not be capable in time to produce a much better successor to this volume, we of the present generation have cause to mourn. For where, more than on the research front, must there be progress! And how can there be progress if students do not excel their teachers—each generation standing on the shoulders of its predecessors, thus fulfilling the aspirations of the generation that begat it.

CHAPTER 2

Library Techniques

CAROLYN W. BOOKWALTER
KARL W. BOOKWALTER

THERE IS AN UNDESIRABLE TENDENCY FOR THE NAIVE OR LESS scholarly student to make his problem outline, or even to collect his data, without first getting an adequate background of the literature available. This usually causes either an unsatisfactory study or difficult and drastic revisions of the written report. Hence, before a person develops a plan for a research problem, he needs to know how the other research in the same area has been done. Not only does he need to know what has been done and how it has been done in the subject area, but he also needs to know the degree of success that was found in the use of the research techniques or methods. The more a person knows about research that has been done and the more he is aware of the gaps and weaknesses of past research, the more apt he is to plan his own research problem well.

Background reading for a research problem should be *completed before* the final plans for the research problem are made. The researcher will thereby benefit from the ideas and knowledge derived from his readings relative to aspects of the problem: ways to select subjects; forms to record and present data; ways to collect, classify, and analyze data; graphic presentation of data; and the form for the writing of the report.

Knowledge of how to locate sources of literature is one basis for high quality scholarship. It saves a person time and effort which could be spent on reading. Knowledge of the sources available in the library should be obtained early in the student's college and university career. He should be constantly alert to additions and changes in the library. Librarians are always glad to help students, but they do not have the time to search for possible specific sources in the literature. This work must be done by the student.

Knowledge of the literature in the field and critical insight into the research in the student's major field of interest is considered evidence of high scholarship. Candidates for higher degrees may not have the experience and background to critically analyze research, but they should work toward that end. Evidence for the principles and methods to be used in new attacks on a research problem will be found in research studies. Students will benefit from time spent in reflective thinking on the readings found in the literature. Reading alone is not enough. The student must become alert to seeing the agreements, differences, and relationships found within and between the sources in the literature. He is then well on the way to becoming a scholar.

PLANNING LIBRARY RESEARCH

Taking time to plan library research will save much time and effort. Time will be saved in the searching and reading processes. The form for entries in the *working* bibliography should be set up. Suggestions for bibliographical form are given in the last chapter of this book, "Writing the Research Report." The working bibliography is that bibliography which includes all possible sources pertinent to the problem. Criteria for a good working bibliography are accuracy, completeness, consistency, and pertinence to the problem.

The working bibliographic card or form for *books* should include the library call number, the author's last name followed by his first and middle names, the title of the book, the city in which the publisher's editorial offices are located, the publisher, the copyright date for a book, and the total pages in the book. The working bibliographic card for *articles* should include the author's last name followed by his first and middle names, the title of the article, the name of the periodical in which the article is found, the volume number, the issue number, the month and year of publication, and the page numbers for the article.

It may be desirable to have two sets of cards of different colors —one for books and one for articles. It may also be desirable to have cards of different colors for the major aspects of the research problem. Or, index cards for major and minor subdivisions might be simpler.

After the form for the working bibliography has been tentatively set up, it is advisable to fill out some cards with several available items, sources of books, articles, and monographs and then see whether the cards can be classified easily, whether all of the necessary information is on the card, and whether there are sufficient spaces for writing the information needed. Suggestions for forms for bibliographic cards are found in *How To Locate Educational Information and Data* by Alexander and Burke (1). Consideration should be given to allowing space on the form for annotations about the value of the sources when reading abstracts of studies or when scanning the sources.

Plans should also be made as to how to take notes when doing the critical reading of sources. An index card box is usually desirable. Notes may be taken on cards, on sheets of paper, or in notebooks. Cards are easier to handle for sorting and reclassifying the notes. Students are cautioned to write on one side of the card, if the cards are to be sorted and classified. If the researcher is sure the bibliography he is using is complete and will not be changed after being put in alphabetical order, the note cards may be numbered to coincide with the number of the alphabetized bibliographical reference. However, most researchers find that they need to change the order of the bibliography, and this changing of the order of a numbered bibliography and the numbered reference notes can become a source of inaccuracy in writing the report. Therefore, a coding scheme for identifying the note card with the correct bibliographical card should be worked out to insure accuracy in the written report. Plans should be made for identifying the kind of notes—quotation, narrative form, and the notetaker's comments or reactions to the material.

This material might be considered as the fact type, the statistical type, the historical type, the how-to-do type, the trends type, or the supporting-evidence type. The researcher may want a combination of these types.

SCOUTING OR SEARCHING

To locate material in the library, the researcher must first know whether needed sources exist. He may go to a number of general sources to find out whether a specific source exists. The accompanying list will help in determining whether a source exists. Once a source is known to exist, then the researcher is faced with the problem of locating the source in a library or in some other location. He should not limit his library research to the sources found in his institution's library. He should cover the area as thoroughly as possible, if a scholarly piece of work is to be done.

DO THEY EXIST?

Articles

- Bulletin Public Affairs Information Service
- Business Education Index
- Business Periodical Index
- Current List of Medical Literature
- Dance Index
- Education Index
- Faculty members and authorities
- Indexes of Journals in special fields not listed in general indexes
- International Index to Periodical Literature
- Index to Legal Periodicals and Law Library Journal
- Music Index
- New York Times Index
- Nineteenth Century Reader's Guide
- Ohio State Periodical Index
- Psychological Index
- Quarterly Cumulative Index Medicus
- Reader's Guide to Periodical Literature
- Research Quarterly Ten-Year Indexes (1930-39, 1940-49)
- Subject Index to Periodicals

Bibliographies

- Bibliography Index
- Bibliographies and Summaries in Education
- Bibliography of Selected Statistical Sources of the American Nations
- Completed Research in Health, Physical Education and Recreation (AAHPER)—annually, beginning in 1959
- International Bibliography of Historical Sciences
- Personnel Bibliography Index
- Research Completed and Research Underway in Health, Physical Education, and Recreation (AAHPER—out of print)
- Annual Bibliographies of Completed Research in Health, Physical Education, and Recreation (AAHPER) 1954-55 (out of print), 1957, 1958
- Research Relating to Children (U. S. Department of Health, Education, and Welfare, Clearinghouse for Research in Child Life)
- Research Studies in Education (Phi Delta Kappa)
- Vertical File Service

Books

- Biography Index
- Cumulative Book Index
- Doctoral Dissertations Accepted by American Universities
- Faculty members and authorities
- Library of Congress Catalogue
- Microcard Bulletin
- National Education Association Handbook, for Local, State, and National Associations
- National Union Catalogue
- Occupational Index
- Publisher's Weekly
- State Law Index
- Subject Index to Reference Books
- Technical Book Review Index
- U. S. Government Printing Office Catalogue

Facts and Statistics

- Accident Facts
- Biennial Survey in Education
- Brief Facts
- Business Statistics

Economic Almanac
Facts on File
Famous First Facts
Information Please
Schoolman's Almanac
Statesmen's Yearbook
Statistical Abstracts of the United States
The World Almanac and Book of Facts

Reviews and Abstracts

Annual Review of Medicine
Biological Abstracts
Book Review Digest
Business Education Index
Child Development Abstracts ...
Bibliographies
Cyclopedia of Education
Dictionary of Sports
Dissertation Abstracts
Encyclopedia Americana
Encyclopedia of American Facts and Dates
Encyclopedia Britannica
Encyclopedia of Educational Research
Encyclopedia of Modern Education
Encyclopedia of Sports
Encyclopedia of Social Sciences
Encyclopedia of Vocational Guidance
Historical Abstracts
Medical Science Abstracts
Mental Measurements Yearbook
Microfilm Abstracts
National Society for the Study of Education Yearbooks
Nutrition Abstracts and Review
Philosophical Review
Physiological Reviews
Psychological Abstracts
Public Health Engineering Abstracts
Review of Data on Research and Research Development Reviews
Review of Economic Studies
Review of Education Research
Review of Social and Economic Research, Health Information Foundations
School Review
Social Science Abstracts
Wistar Institute of Anatomy and Biology Abstracts
Yearbook of Education

ARE THEY IN THE LIBRARY?

Card Catalogues—Author, Subject
Library Departmental Catalogues (Circulation, Documents, Periodicals, Reference, Reserve)
Reference Librarian
Libraries of Faculty Members or Authorities in the Field
Special Collections

INDEXES

General indexes such as *Education Index* and *Reader's Guide to Periodical Literature* index only those periodicals listed at the front of the volume of the index. For instance, the *Journal of Health-Physical Education-Recreation* is listed in *Education Index* but *The Physical Educator* is not presently listed. It is wise to check the indexes of those periodicals which are not listed in the general indexes for possible articles related to the problem undertaken.

Education Index is a comprehensive author and subject guide to educational literature from all significant American sources and a few British sources since January 1, 1929. Over 150 educational periodicals and yearbooks are indexed. Those periodicals pertinent to the fields of health, physical education, and recreation which are listed in *Education Index* are marked with an "E" in the list of periodicals below. All educational books published in the United States (including college textbooks but not elementary and secondary school textbooks) are indexed in this source. Practically all of the publications of the United States Department of Health, Education, and Welfare and of the National Education Association are included. Monthly supplements are issued from September through May. Clothbound, three-year cumulative volumes were issued from 1929 to 1953. Beginning in 1954, each even-numbered year has a one-year cumulation, with two-year cumulations being issued on odd-numbered years.

For coverage before 1929, the researcher will have to go to sources such as *International Index to Periodical Literature*, the *Ohio File*, and the *Record of Current Educational Publications* for the years 1920-28. For the years 1912-19, the coverage of sources will be found in *Reader's Guide Supplement*, as well as in the *Record of Current Educational Publications* and the *International Index to Periodical Literature*. For the years 1907-11, one may find sources in *Reader's Guide* and *Reader's Guide Supplement*. Before 1907, coverage may be found in *Reader's Guide, Annual Library Index, Poole's Index to Periodical Literature*, and *Nineteenth Century Reader's Guide to Periodical Literature*.

Reader's Guide to Periodical Literature is an author and subject index of articles of popular and general nature from over 130 magazines. It rarely duplicates the indexing in the *Education Index*. It has been published since 1900. Before 1929, it covered many journals in the field of education which were transferred to *Education Index* in 1929. For a coverage of articles before 1900, the researcher would have to go to *Poole's Index to Periodical Literature* and *Nineteenth Century Reader's Guide to Periodical Literature*. *Reader's Guide*, published semi-monthly from September to June and monthly in July and August, is bound in a cumulative annual index. Those periodicals pertinent to health, physical education, and recreation which are indexed in the *Reader's Guide* are marked with an "R" in the list of periodicals below.

Bibliographic Index is a source of bibliographies, including mimeographed and multilithed bibliographies in a wide range of subjects. It was first published in 1938. It is published quarterly and is bound as a cumulative annual index and a four-year index.

Bibliography Index was started in 1946. About 1500 periodicals are examined regularly for bibliographical material and indexed. The index is published in November, February, May, and August and is bound as a cumulative index annually.

Book Review Digest contains about 4000 book reviews yearly. The book reviews are listed by author and the cumulative subject and title indexes are alphabetically arranged in the back of each bound volume. Both favorable and unfavorable reviews are given. It was first published in 1905 and is now published monthly except in July. There are six-month, yearly, and five-year cumulative indexes.

Business Periodical Index is a source for about 120 periodicals in the fields of accounting, advertising, banking and finances, general business insurance, labor and management, marketing and purchasing, office management, public administration, taxation, specific businesses, industries, and trades. It was started in January 1958 and is published monthly.

Cumulative Book Index was started in 1898 as the *U. S. Catalogue*. It indexes books published throughout the world in the English language. It is indexed by author, editor, subject, and translator. Included in the information provided are the publisher, price, date of publication, paging, size, edition, and Library of Congress order number. It appears annually as well as monthly.

Current List of Medical Literature was begun in 1941 as a monthly and cumulative index. Since 1945 it has been published weekly, with a monthly subject index and an annual cumulative index. About 1200 medical journals are indexed by subject. It is published by the Army Medical Library.

Filmstrip Guide is published semi-annually. Lists of 35mm filmstrips are indexed by title and subject and classified by the Dewey Decimal System. Annual supplements also index 16mm films. There is an annual cumulative index.

International Index to Periodical Literature contains an index of about 250 periodicals related to pure science and the humanities. From 1920-28, technical and specialized articles in education were indexed, but in 1929 these were transferred to *Education Index*. The *International Index* was started in 1920 and is published monthly. There are cumulative quarterly and three-year indexes.

Index to The Times (London) was started in 1946. From 1946 to 1957, it was published quarterly. Since that time, it has been published every two months.

Library of Congress Catalogue of Motion Pictures and Filmstrips is published annually. It is indexed by subject.

The New York Times Index is a source of news items listed by subject. The date, page, and column in which the material is found in *The New York Times* is given. The index appears annually.

Psychological Index is the only large comprehensive guide to psychological literature for the years 1894 to 1935. Books and periodicals in all languages are indexed. Publication ceased in 1935.

Quarterly Cumulative Index Medicus has been in existence since 1927. It covers about 1200 periodicals, including foreign ones. It has an author and a subject index of writings in medicine and related sciences. It is published semi-annually, each index appearing about two years after the publication of the periodicals.

Reference Catalogue to Current Literature is a reference index of books printed in the British Empire. The books are indexed by author and by title. Details given are author, title, editor, translator, reviser year of publication, number of editions, size, series and binding, number of pages, and illustrations and illustrator. It is published annually.

Subject Index to Periodicals is published quarterly in Great Britain. About 300 periodicals are indexed. There is an annual cumulative index.

In addition to the indexes and general reference materials listed and/or described above, there are some rather specific sources which will aid the researcher in locating other sources.

Specific Sources

AFFLECK, GEORGE B. "Bibliographies." *American Physical Education Review*, March 10, 1910—June 1929; *Research Quarterly*, December issue, 1930-41.

AMERICAN ACADEMY OF PHYSICAL EDUCATION. *Professional Contributions*, Nos. 1-6. Washington, D. C.: the American Association for Health, Physical Education, and Recreation, a department of the National Education Association, November 1951-58.

AMERICAN ASSOCIATION FOR HEALTH, PHYSICAL EDUCATION, and RECREATION. *Annual Bibliography of Completed Research*. No. 1, 1956; No. 2, 1957; No. 3, 1958. Washington, D. C.: the Association, a department of the National Education Association.

AMERICAN HISTORICAL ASSOCIATION. *List of Doctoral Dissertations in History Now in Progress at Universities in the United States*. Washington, D. C.: the Association, Library of Congress Annex, 1958, 64 p. (Revised every three years).

AMERICAN JOURNAL OF SOCIOLOGY. "A Doctoral Dissertation in Sociology, 1952." *American Journal of Sociology* 59: 76-85; July 1953.

BROWN, STANLEY B.; LYDA, MARY LOUISE; and GOOD, CARTER V. *Research Studies in Education, A Subject Index.* Bloomington, Ind.: Phi Delta Kappa, Annual Supplement.

BURCHFIELD, LAVERNE. *Student's Guide to Materials in Political Science.* New York: Henry Holt and Co., 1935. 426 p.

BURKE, A. J., and ALEXANDER, CARTER. "Guide to the Literature on Public School Administration." *Elementary School Journal* 37: 764-78; 1937.

CONLEY, WILLIAM H., and BERTALEN, FRANK J. *Significant Literature of the Junior College, 1941-1948.* Washington, D. C.: American Association of Junior Colleges, 1949. 40 p.

CONARD, RICHARD. "Systematic Analysis of Current Researches in the Sociology of Education." *American Sociological Review* 17: 350-55; June 1952.

COOK, W. H. "A Guide to the Literature on Negro Education." *Teachers College Record* 34: 671-77; May 1933.

COOPER, ISABELLA M. *Bibliography on Education Broadcasting.* Chicago: University of Chicago Press, 1942. 576 p.

COULTER, EDITH M., and GERSTENFELD, MELANIE. *Historical Bibliographies.* Berkeley: University of California Press, 1935. 206 p.

CURETON, T. K. *Masters' Theses in Health, Physical Education, and Recreation.* Washington, D. C.: American Association for Health, Physical Education, and Recreation, a department of the National Education Association, 1952. 292 p.

DECK, A. E. "A Guide to the Literature of the Curriculum." *Teachers College Record* 35: 407-14; February 1934.

DOMAS, SIMON J., and TIEDEMAN, DAVID V. "Teacher Competence: An Annotated Bibliography." *Journal of Experimental Education* 19: 101-218; December 1950.

FRAZIER, B. W. *Education of Teachers, 1935-1941.* Office of Education, Federal Security Agency, Bulletin 1941, No. 2. Washington, D. C.: Superintendent of Documents, Government Printing Office, 1941. 60 p.

GEORGE PEABODY COLLEGE FOR TEACHERS. *A Survey of Surveys.* Nashville, Tenn.: the College, Division of Surveys and Field Services, 1952. 52 p.

GOHEEN, HOWARD W., and KAVRUCK, SAMUEL. *Selected Reference on Test Construction, Mental Test Theory, and Statistics.* Washington, D. C.: Superintendent of Documents, Government Printing Office, 1950. 209 p.

GOOD, CARTER, and *others.* "Summary of Studies Relating to Exceptional Children." *Journal of Exceptional Children.* Extra issue. January 1938. 60 p.

GOODENOUGH, FLORENCE L. "Bibliography in Child Development 1931-1943." *Psychological Bulletin* 41: 615-33; November 1944.

KAPLAN, LOUIS. *Research Materials in Social Sciences.* Madison: University of Wisconsin Press, 1939. 36 p.

LAMKE, T. A., and SILVEY, HERBERT M. *Masters' Theses in Education, 1951-52.* Cedar Falls, Iowa: Research Publications, 1953. 168 p.

LAYTON, ELIZABETH N. *Surveys of Higher Education in the United States, 1937-1949.* U. S. Office of Education, Federal Security Agency. Washington, D. C.: Superintendent of Documents, Government Printing Office, May 1949. 24 p.

LOUTTIT, C. M. *Bibliography of Bibliographies on Psychology, 1900-1927.* Washington, D. C.: National Research Council, 1928. 108 p.

LYDA, MARY L., and BROWN, STANLEY B. *Research Studies in Education 1941-1951.* Boulder, Colorado: the Authors, 1953.

MCGRATH, EARL J. *Bibliography in General Education.* Washington, D. C.: American Association of Junior Colleges, 1950. 32 p.

MCSWAIN, E. T., and ALEXANDER, CARTER. "Guide to the Literature on Elementary Education." *Elementary School Journal* 35: 747-59; June 1935.

MANSKE, A. J., and ALEXANDER, CARTER. "Guide to the Literature on Secondary Education." *School Review* 42: 368-81; May 1934.

MAY, MARK A., and DOOB, LEONARD. "Research on Competition and Cooperation." *Sociological Science Research Council Bulletin, #25.* New York: the Council, 1937. 200 p.
PALFREY, T. R., and COLEMAN, H. E. *Guide to Bibliographies of Theses—U. S. and Canada.* Second edition. Chicago: American Library Association, 1940. 54 p.
SEGEL, DAVID. "Education Research Studies of National Scope or Significance." *Biennial Survey of Education in the United States, 1938-40.* U. S. Office of Education, Federal Security Agency. Washington, D. C.: Superintendent of Documents, Government Printing Office, 1941. Vol. 1, Chapter 10.
SMITH, A. H. *A Bibliography of Canadian Education.* Department of Educational Research, Bulletin No. 10. Toronto: University of Toronto, 1938. 302 p.
SMITH, H. L., and PAINTER, W. T. "Bibliography of Literature on Education in Countries Other than the United States of America." (1919-24) *Bulletin of the School of Education.* Bloomington: Indiana University, 1938. Vol. 14, No. 3. 144 p.
UNESCO. *Handbook of Educational Organization and Statistics.* New York: Columbia University Press. Every three years, 1951-.
U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, CHILDREN'S BUREAU. *An Evaluative Study of Research in School Health Services.* Washington, D. C.: Superintendent of Documents, Government Printing Office, 1958.
WALKER, GEORGE H. "Masters' Theses Underway in Negro Colleges and Universities, 1951-52." *Negro Educational Review* 3: 68-79; April 1952.

Faculty members and authorities in the field may be able to suggest sources of materials not indexed in general indexes or journals. Such materials might consist of papers presented at association meetings or materials distributed by professional groups or conferences.

Bibliographies of books, theses, and research papers are valuable sources for locating the existence of pertinent literature. The bibliographies in the *Review of Educational Research* and *Encyclopedia of Educational Research* should not be overlooked. Sometimes bibliographies may be found in periodicals. For example, the "Affleck Bibliography" was published annually in the *American Physical Education Review* from 1911 to 1929, and in the *Research Quarterly* from 1930 to 1941. In the March 1949 issue of the *Research Quarterly* will be found an excellent bibliography of doctoral theses completed from 1930 to 1946—bibliography compiled by T. K. Cureton. There are many other bibliographies to be found in the *Research Quarterly,* especially in the issues from 1930-40, 1951, and 1952.

Some confusion may arise from the change in titles of periodicals. For example, *Today's Health* was formerly *Hygeia.* However, the cross reference on the library card catalogue will usually refer to the former titles.

The periodicals listed in the following group are not all of the periodicals which may have material of value to various aspects of the fields of health, physical education, and recreation. Publications of the organizations affiliated with the American Association for Health, Physical Education, and Recreation should be checked for possible sources. In addition, there are book and article reviews or abstracts in many of the periodicals listed below. For example, the *Research Quarterly* has research abstracts and the *Physical Educator* has book reviews and article reviews. Those journals marked with "E" are listed in *Education Index,* those marked "I" are listed in *International Index to Periodical Literature,* and those marked "R" are listed in *Reader's Guide to Periodical Literature.*

Periodicals

E Adult Education
E American Association of Colleges for Teacher Education Yearbook
E American Business Education
E American Childhood
American Journal of Anatomy
American Journal of Clinical Nutrition
R American Journal of Hygiene
American Journal of Occupational Therapy
American Journal of Physical Medicine
American Journal of Physiology
E American Journal of Physics
American Journal of Psychiatry
American Journal of Psychology
American Journal of Public Health and the Nation's Health
I American Journal of Sociology
American Public Health Association Yearbook
American Review of Tuberculosis
E American School Board Journal
I American Schools of Oriental Research
I American Sociological Review
American Statistician
E American Teacher Magazine
Annual Review of Medicine
Annual Review of Physiology
R Archives of Physical Medicine and Rehabilitation
E Association for American College Bulletin
Association for Physical and Mental Rehabilitation Journal
E Association of School Business Officials
E Athletic Journal
Audubon Magazine
Baseball Magazine
Better Schools
Biological Bulletin
E British Journal of Educational Psychology
British Journal of Nutrition
British Journal of Psychology
Bulletin of Hygiene (London)
E Bulletin of the National Association of Secondary School Principals
E Business Education Forum (UBEA)
Business Statistics
California University Folklore Studies
E Camping
Canadian Public Health Journal
Cancer Research
E Child Development
E Childhood Education
E Children
E Child Study
E Clearinghouse
Coach and Athlete
College and Research Libraries
Comparative Education Review
R Congressional Digest
Congressional Index
R Current History
Current Medical Digest
Current Sociology
R Dance Magazine
Dance Observer
Economic Digest

E Education
E Education Administration and Supervision
E Educational and Psychological Measurements
Educational Film Guide
Educational Focus
E Educational Leadership
E Educational Research Bulletin
E Education Digest
Family Life
Film News
Filmstrip Guide
Flying Safety
Folk Dancer
Good Health
Health Instruction Yearbook
Hearing News
E Higher Education
Historical Journal
Historical Studies
History of Education Journal
Immunity Bulletin
E Indiana University School of Education Bulletin
Industrial Hygiene Bulletin
Institute of Historical Research Bulletin
Institute of International Education News Bulletin
International Journal for Health Education of the Public
International Journal of Group Psychotherapy
I Journal of American Folklore
Journal of Applied Physiology
Journal of Applied Psychology
E Journal of Business Education
Journal of Child Psychiatry
E Journal of Counseling Psychology
E Journal of Consulting Psychology
Journal of Correctional Education
Journal of Diseases of Childhood
E Journal of Education
E Journal of Educational Psychology
E Journal of Educational Research
E Journal of Educational Sociology
ER Journal of Experimental Education
E Journal of General Education
Journal of General Physiology
Journal of General Psychology
Journal of Genetic Psychology
E Journal of Health-Physical Education-Recreation
E Journal of Higher Education
Journal of History of Ideas
E Journal of Negro Education
Journal of Nutrition
Journal of Personality
Journal of Physical Education (YMCA)
Journal of Psychology
E Journal of School Health
Journal of Social Psychology
E Journal of Teacher Education
Library Review
Mental Hygiene
Mentor
Mind, A Quarterly Review
Modern Humanities Research Association Bulletin
Modern Schoolman
Monthly Bulletin of Statistics
Monthly Catalogue of Government Publications
Monthly Checklist of State Publications
Monthly List of Books Catalogued in the Library of the United Nations
National Association of Business Teacher Education Bulletin
National Bureau of Economic Research
E National Business Education Quarterly
ER National Education Association Journal
E National Education Association Research Bulletin
E National Elementary Principal
R National Institute of Health Bulletin
National Mental Health Program—Progress Report
National Negro Health News
E Nation's Schools
Negro Education Review
New Publications of the United Nations Headquarters
Occupational Psychology
Occupational Safety and Health
Ohio State University Education Research Bulletin
R Outdoor Life
E Peabody Journal of Education
E Phi Delta Kappan
Physical Educator
Physical Therapy
Population Index
Progressive Education
Psychological Bulletin
Psychological Monographs
Psychological Reports
Psychometrika
Public Health Nursing

Public Health Reports
Public Safety
Public Welfare
Publicity Problems
Publisher's Circular
Quality Control and Applied Statistics
Quarterly Bulletin of Fundamental Education
Quarterly Journal of Experimental Physiology
Quarterly Journal of Experimental Psychology
Quarterly Journal of Studies in Alcohol
R Reader's Digest
R Recreation
E Religious Education
E Research Quarterly (AAHPER)
Research Reviews
Research Today
I Review of Economic Studies
E Review of Educational Research
E Safety Education
Safety Standards
E Scholastic Coach
Scholastic Teachers
School Activities
ER School and Society
E School Executive
ER School Life
E School Review
R Science Digest
Social Research
Social Science Research Council Bulletin
E Social Studies
Society for Research in Child Development Monograph
I Sociological Review
Sociometry
I Spectator
Sporting News
Sports Illustrated
Stanford Research Institute Journal
Statistical Bulletin
Student Life
Swimming Pool Age
Teacher Education Quarterly
Teaching Tools
Textile Research Journal
R Today's Health
Universities Review
Weekly Review of Periodicals
Yearbook of Education
Youth Leaders' Digest

In searching indexes, the researcher may have to look under various subject headings other than the title of his research problem. Related topics should be investigated in addition to the specific topics. The following exemplary list will give the researcher an idea of some of the possibilities of indexed topics.

Ability, also refer to Performance and Achievement.
Achievement, also refer to Ability and Performance.
Achievement Tests, a major heading under Health Education, Physical Education.
Activities, also refer to Sports, Games, Recreation.
Activity Tests, also refer to specific activities under Achievement Tests.
Addresses of Periodicals and Publishers, see Periodicals, addresses, and Publishers, addresses.
Age, also refer to Chronological Age, Physiological Age, Anatomical Age, Classification Bases.
Appraisal, also refer to Aptitude, Character, Personality, Body Type, Evaluation.
Attitude, refer to Behavior, Sportsmanship, Character, Personality.
Bibliography, see particular subject for subhead. Also see Booklists.
Body Build, also refer to Constitutional Type, Body Type, Physique.

Body Mechanics, also refer to Posture, Flexibility, Joints, Physical Proportions, Anthropometry, Kinesiology.
Book Reviews, main heading. Look for author or subject under heading.
Buildings, also refer to Gymnasium, Pool, Fieldhouse, Track, Playground, Courts, Equipment.
Character, refer to Tests, Appraisal, Personality Traits.
Circulatory, also refer to Functional Tests, Respiratory-circulatory or Circulatory-respiratory Tests, Cardiac, Cardiovascular Tests.
Condition, also refer to Physical Fitness, Training, Physique, Functional Tests.
Constitutional Type, also refer to Body Type, Physique, Functional Tests.
Co-ordination, also refer to Motor Ability, Motor Fitness, Agility Balance, Skill. Also see specific activities.
Correlation, refer to Statistical Research Methods.
Development, refer to Growth, Physical Development, Maturation.
Education Research, refer to Research and see subheads.
Endurance, refer to Stamina, Organic Fitness, Muscular Endurance.
Environmental Conditions, refer to Temperature, Barometric Pressure, Relative Humidity, Noise, Lighting.
Equipment, refer to Athletic Equipment, Supplies, Uniforms, Balls, Nets, Racquets, Apparatus, Aquatic Equipment, etc.
Factor Analysis, refer to Statistical Research Methods.
Fat, refer to Adipose Tissue, Endomorphic.
Federal Documents, see Government Documents.
Films, refer to Cinematography, Moving Pictures.
Fitness, refer to Physical Fitness, Health, Measurement.
Flexibility, refer to Suppleness, Litheness, Joint Movements.
Girth, see subheads Growth, Body Build, Nutrition, Physique, and Condition Tests.
Health, refer to Physique and Condition Tests.
Interest, refer to Behavior, Sportsmanship, Character, and Personality Type Tests.
Marking, refer to Administration of Tests and Testing Procedures, Grades.
Maturation, refer to Growth Development.
Medical Examination, refer to Physique and Condition Tests, Health Examination.
Motor Ability, refer to Achievement Tests.
Muscle, refer to Physique and Condition, Strength, and Power.
Nutrition, refer to Physique and Condition Tests, Health Tests.
Objectivity, refer to Administration of Tests and Testing Procedures.
Percentiles, refer to Statistical Procedures.
Performance, see Achievement, Ability.
Physical Fitness, refer to Fitness, Physical Education, Health, Physical Condition.

Power, refer to Physique and Condition, Achievement Tests.
Prediction, refer to Administration of Tests and Testing Procedures Also see Statistical Procedures.
Puberty, refer to Maturity, Adolescence.
Scales, refer to Scales, Norms.
Skill, refer to Achievement Tests and to various sports.
Standards, refer to Administration of Tests and Testing Procedures, Norms, Score Cards.
Strength, refer to Physique and Condition Tests.
Subject headings, often under subhead Bibliography. Also see Book-lists.
Surveys, see subject heading, e.g., Health.
Vital Capacity, refer to Physique and Condition Tests, Lung Capacity.
Weight, see Physique and Condition Tests, Growth, Nutrition.

LOCATING THE SOURCES

When the researcher has established a bibliography of existing sources, he then needs to locate the sources. The card catalogue in a library is the inventory of the library. If the author is known, usually the simplest procedure is to use the author index. From the author index, the correct title of the desired sources may be obtained and sometimes additional sources by the author may be found. The call number and occasionally the specific location of the source may be obtained from the author card. If the author is not known, the subject index should be used. This may entail looking under several different headings, as indicated above. Subject cards will frequently give cross-references to other subjects. The most complete card, as a rule, is that under the author.

Sources in libraries are classified by one of two systems. The Dewey Decimal System is usually used in small libraries. The numbers range from 000 to 999. The areas in which researchers in health, physical education, and recreation will find sources are references—000, philosophy—100, sociology—300 [under which statistics is 310, political science is 320, law is 340, administration is 350, associations and institutions are 360, education is 370 (including health, physical education, and recreation)], fine arts—700, literature—800, and history—900.

The Library of Congress System is used by large libraries. The symbols range from A to Z and combinations of the letters of the alphabet. The areas in which researchers in health, physical education, and recreation will find sources are: A—general works;

B—philosophy-religion; C—geography-anthropology; G—geography-anthropometry (Physical education and recreation are found under GV.); H—sociology; K—law; L—education (Theory books on education including many theory books in health education, physical education, and recreation education will be classed under LB, and education dissertations will be found classed under LZ.); M—Music; N—fine arts; R—medicine (including health areas and hygiene); U—military science; V—naval science; and E—bibliography and library science.

If the library does not have the desired sources, there are several possibilities for obtaining them. Some sources may be available from the publishers. When it is not possible or feasible to obtain them from the publisher, the Inter-Library Loan Service may be used for obtaining sources located in some other library. Sources are generally loaned for a period of two weeks. The borrower may have to pay the transportation charges or a fraction of them. Some researchers have found it more desirable to purchase a microfilm or microcards of the source than to pay the transportation charges on the original source. There are, of course, certain sources that are not obtainable on loan. Periodicals and rare volumes come under this category.

SCANNING OR SKIMMING

As one is able to locate a source, he needs to evaluate the source for its potential value to the solution of the research problem. One should skim the literature for each major purpose in the problem. Titles of books and articles can be misleading at times. The copyright date of a book or the date of publication of the article should be noted. By starting with the more recent literature when taking notes, references will be found to older sources. Likewise, it is well to start with more general sources and thereby obtain insight as to the likely specific sources. Sometimes the source from which a bibliographical reference was obtained was erroneously printed. The bibliographic card should be checked to be sure all data are correct.

By scanning the table of contents or the index, the researcher may determine whether there are topics on the desired subject included in the book. By scanning the center headings and summaries of chapters, more information may be gained as to the probable value of the source for data.

If annotations are made on the working bibliography cards as to the values of the sources, the researcher will have a base for selecting pertinent sources. The annotations will also guide the researcher as to his emphasis in notetaking. Reading available abstracts will aid the researcher in verifying his own estimate of the source. The annotation of the source should include the organization of the material, the number and quality of illustrations, the emphasis in the source, the research techniques or methods used, and the essential findings. Suggested forms for annotations may be found in *How To Locate Educational Information and Data* by Alexander and Burke (1).

CRITICAL READING OR GLEANING

It is well to read the summary of the research article or of the chapter of the book before reading the contents. The reading of the summary gives the reader a total picture of the chapter or article and of the things which the author considered to be important. By noting the center heads and side heads of the literature, the reader may have a fairly accurate outline of the material.

When reading, the researcher should try to have certain questions in mind for which he is trying to find the answers. This will aid him in concentrating upon the subject.

While doing critical reading, the researcher will want to take notes. If the note-taking form has been well planned, provision will be made for the researcher to add his own comments and ideas which occur as he skims or gleans the reference materials. It is better to err on the side of taking too many notes rather than not enough notes. Criteria for good notes are completeness (in scope), ease of assembly (flexibility), expansibility, uniformity (consistency), and pertinence (essential to the problem). Legibility and careful identification of the notes will improve their accuracy. Quoted materials should be quoted *accurately*. Even misspelled words should be spelled the way they were found, but the researcher should indicate that the error occurred in the quotation and was not an error in note-taking. It is best to identify quotations with quotation marks, so that several weeks or months later there is no question as to whether the material is a quotation or an interpretation. Where one source quotes another source, the researcher should attempt to find the primary source and obtain the quotation from it. The researcher will then be assured that the material is quoted accurately.

When making narrative notes, care should be taken that material is not quoted without giving due credit for the quotation. Bibliographical reference should be given for ideas obtained from sources. Credit should be given for unique words or phrases.

Notes in outline form are used only by the experienced researcher who is looking mainly for ideas. Even in this form of note-taking, bibliographical reference should be given to the notes.

After notes have been taken from a few sources, it is worthwhile to make a pilot study of the notes. Sort, classify, and try to analyze the notes. After the pilot study, the researcher will be able to correct any faults in note-taking while it is feasible to do so. He will also obtain insight into the values of the sources.

After all sources have been explored and notes completed, the researcher should then spend some time in processing and analyzing his notes. Insight into agreements, differences, relationships, and trends will be obtained by the reflective thinking given to the results of the analyses of the notes. Quality, even more than quantity, of ideas discovered in the literature is needed for good research. Too many graduate students merely write up one abstract of literature source after another and do nothing about analyzing the material. The researcher should determine the agreements and differences among the sources and between the sources and the research problem being undertaken. This process will involve reflective thinking and careful writing.

The material should be written up in the best form possible. The final chapter in this book, "Writing the Research Report" will be helpful in the writing process.

The better the library research, the better will the researcher be able to carry out his selected problem. The reading should be done *before* the selected problem has been outlined and the research design completed, if the researcher is to do a scholarly piece of research. As a result of broad and deep reading, the researcher will have obtained ideas on the phases of the problem, the values of certain methods and techniques, the possibilities of analysis of data, and ways of presenting the analysis of data. These ideas will help the researcher focus clearly upon the various aspects of his own problem, and he will have a background upon which to build research designs.

BIBLIOGRAPHY

1. ALEXANDER, CARTER, and BURKE, ARVID J. *How To Locate Educational Information and Data.* Fourth edition, revised. New York: Bureau of Publications, Teachers College, Columbia University, 1958. 419 p.
2. AMERICAN ASSOCIATION FOR HEALTH, PHYSICAL EDUCATION, AND RECREATION. *Research Methods Applied to Health, Physical Education, and Recreation.* First Edition. Washington, D. C.: the Association, a department of the National Education Association, 1949. p. 84-124.
3. GOOD, CARTER; BARR, A. S.; and SCATES, DOUGLAS E. *The Methodology of Educational Research.* New York: D. Appleton-Century Co., 1941. p. 104-84, 856-63.
4. GOOD, CARTER, and SCATES, DOUGLAS. *Methods of Research.* New York: Appleton-Century-Crofts, 1954. p. 133-69.
5. LARSON, LEONARD A.; FIELDS, MOREY R.; and GABRIELSON, MILTON A. *Problems in Health, Physical Education, and Recreation Education.* New York: Prentice-Hall, 1953. p. 179-92.
6. WILLIAMS, CECIL B., and STEVENSON, ALLAN H. *A Research Manual.* New York: Harper and Brothers, 1951. p. 16-61, 85-95.

CHAPTER 3

Selection and Definition of a Problem

KARL W. BOOKWALTER

BY FAR THE GREATEST MAJORITY OF STUDENTS IN THE AREAS OF health, physical education, and recreation come to their graduate work without any idea of what their problem for their required thesis or dissertation might be. This is probably a blessing in disguise. Most people lack the curiosity, insight, and techniques that make them problem-conscious as they enter upon their graduate work. The seeker for a problem may be assured that he is not peculiar or unique. He has company. In fact, this need is probably the basis for the usual requirement of a course in the Introduction to Research.

Other characteristics of the average problem seeker are to grasp at "the first straw"—the first problem that suggests itself or is suggested by his adviser or by some professor. Again, and probably more typical, is the tendency for the student to try to work out some vague outline of a problem out of his experience or imagination, practically unaided by background reading, and to hasten to have the concept reserved or accepted at once. This leads to the downfall of many of the students who have completed all requirements for a degree except the thesis. These students comprise the army of "course takers" who drop out because of failure to find or to complete the required problem.

A problem properly chosen and carefully planned is practically assured of success for the average or better than average doctoral candidate.

This chapter is included in this book because the selection and definition of a problem is a major obstacle to the success of many researchers, because a proper start is so vital, and because the student can be helped in this quest.

HOW TO LOCATE A PROBLEM

The student may turn to his own experience, to his educational background, and to direct searching in order to locate a problem.

Experience in His Vocation. As a person teaches health or physical education for his first year or so, he is concerned with the mechanics, with finding proper sources, with proper sequences, and with proper time allotments. Methods too must be mastered. Little time exists in this trial and error process to note, much less formulate and execute, problems which confront all or many in this field. Then come the years of curiosity, of relaxed control, and of time to tackle practical problems more slowly and more systematically and carefully.

Coaching, administration, or the conduct of recreation programs follow the same pattern—first confusion, then control, then curiosity. This last trait, unfortunately, is possessed by too few and in too small a degree. Ten years of experience routinely following in the same rut is only one year's experience repeated ten times. A person must be curious, alert, creative, and capable if he is to contribute to his profession by joining in its problem solving.

It is in this crucible of practical experience that the recognition of problems and the depth of understanding regarding probable solutions, and related and dependent variables necessary to problem solving, are acquired. It has been said that the techniques of research can better be given to the field expert than can field expertness be imparted to the research technician.

If the researcher's experience includes use and manipulation of equipment, stadiometers, sphygmomanometers, spirometers, goniometers, and such, likely problems and appropriate techniques present themselves readily. When he is a stranger to such technical equipment, then certain areas of research interest are probably closed to him. If tests and measurements of pencil and paper or performance type pique his interest, he sees certain problems of which a more disinterested person is unaware.

In short, the richer the researcher's experience, the more the likelihood of his being constantly confronted with problems which he might like to tackle. The more narrow or mechanical his professional experience, the greater the probability that problem solving is not for him. At least the task will be difficult and somewhat of a chore.

Contacts with the fields of physiology, physics, psychology, kinesiology, or other laboratory work, may stimulate curiosity and research for the person in the areas of health, physical education, and recreation. These fields are uniquely tied to the laboratory or experimental approach to problems. A person with such contacts might well be expected to find his *problem* or problems in this area.

Familiarity with data frequently found in safety divisions, in health agencies, in guidance departments, and in welfare and youth-serving agencies should make a person extremely curious. Are the records or tests valid, reliable, or objective? Could a better test be made for the purpose? What are the interrelationships? What makes for failures? What for success? Only the most routine and clerical mind would fail to want to know the answers to these and myriads of similar questions. Records are collected essentially to aid in solving the organization's problems. Only when they are thus used is the time spent in obtaining these facts justified.

In short, if a person is *not* puzzled, curious, or challenged, maybe he should not go on for higher degrees. On the other hand, these curiosities can, to some extent, be stimulated by study and vicarious or institutional experiences.

Education and Training. A student's professional subjects will of course be pointed toward his major goal in life. However, that goal today should, on the graduate level, include an interest in solving or helping solve the problems in his field.

Technical skills commonly required for research are statistical organization, analysis and interpretation, experimental design, measurement, survey techniques, documentary analysis, action research, and many others. Regardless of his field, the *leader* will need to be reasonably well equipped in many of these.

A course in introduction to research is basic to an overview of these potentialities. This chapter is especially pointed to this end. The techniques suggested in this and other chapters are, as indicated, introductory only. If, for example, a student is to do a statistical, experimental, or measurement study, from six to nine hours of statistics are a basic requirement. Otherwise the study should not be undertaken. Similarly, historical, physiological, psychological, or kinesiological studies should not be undertaken by the naive person. There are no short cuts to a good study and certainly there are none to continued contributions to one's field.

In any graduate course—such as the Organization and Administration of Health Education, or of Physical Education, or of Recreation, or in any good Principles course—the student should be confronted with unanswered problems, authorities in such fields, and resource materials which might be helpful. Seminars should stimulate his desire to learn and his creative thinking along certain lines, if he is truly a graduate student. His curiosity and creativity may be stimulated by the faculty members teaching such advanced courses.

These things do not come all at one time. They are part and parcel of a good graduate education. The student can be too quick to decide what he will choose for his initial research problem. On the other hand, many degrees are lost by procrastination. The student should begin thinking creatively early in his graduate career. Then, when he is sure and capable, he can tackle the problem.

Direct Search for a Problem. Talks with his professors and with those in related fields may reveal to the student less personal problems which may pique workers and scholars. Researchers frequently add to the unsolved areas in the very process of having solved certain problems. Though the best problem for an individual is his own problem, investigation of the problems of others may lead to a vicarious interest which will transform the searcher into a researcher. A warning is in order that such a step may lead only to a blind alley or at best to a boresome experience solely to meet a requirement. On the happier side, it might be recalled that earlier custom prescribed that the hapless candidate do the problem assigned by his professor. So, it *has* been done with success by others in the past.

The hard way to search for a problem is for the student to set himself to the task of laborious library scouting and scanning. Since there is an element of aimlessness about such reading, or seeking, at first it is likely to be in great part ineffectual. However, once the student is on the track of a likely topic, this experience will serve him in gor l stead. Many sources and voluminous notes should be on hand tart him on his particular project. He will have acquired a degree of critical skill in library work already. The preceding chapter on library techniques will be helpful in this approach.

CRITERIA FOR AN ACCEPTABLE PROBLEM

As the researcher gathers experience on the job and is enlightened by his graduate education and by much searching, scanning, and skimming in the library, it is quite likely that many or several potential problems will present themselves. Rather than a dearth, there may suddenly be a plethora of ideas. To jump at the first one or to select one at random would be foolhardy. The right choice is extremely important. While the choice of a problem is the responsibility of the candidate, there are nevertheless some bases by which a more intelligent decision may be made. These criteria are both personal and social in nature.

The usual first research project is unlikely to be epoch making. So, self-preservation being nature's first law, the researcher naturally turns to self-analysis in preliminary appraisals of potential problems.

First, he must be sure the topic is *definitely and specifically delimited* so that its scope and difficulty are accurately foreseen. The basic assumptions must be known and it must be evidenced that they can be met. The hypotheses must be known and be so stated that in the end their tenability or verity can be established or rejected logically and/or statistically. Usually a problem first presents itself in somewhat nebulous form. Also, its scope tends to be such that it is impossible of attainment by one person, at least in a reasonable time. Proper definition and planning can usually reduce the idea to manageable form and its scope to attainable proportions. Proper definition of terms will aid in both efforts. When the student is reasonably assured that there are merits in the proposed topic, complete analysis by the horizontal analysis technique, explained and exemplified below, will practically assure him that he can answer this first criterion and those following.

Just for example, the idea may come to him, "What are the effects of athletics?" Obviously he would need to decide: "upon boys or girls in elementary school, junior high school, senior high school, or college?" Having decided the problem will be "What are the effects of athletics upon junior high school boys?," he starts out on the quest for certainty. Soon he realizes that there are degrees of athletics and a variety of sports. Again he has to decide upon the more vigorous interschool sport and the one with most debatable safety, football. Now the problem is "What are the effects of interscholastic football upon junior high school boys?" Pointed reading now becomes more possible and also more enlightening. He soon discovers that there are myriads of possible effects—scholarship, school tenure, physical fitness, growth and development, injuries, mental health or adjustment, acceptance by the group, or even juvenile delinquency.

For the point of argument, let us assume that he has recently heard of the Wetzel Grid (23) in a psychology class or in tests and measurements (or both). This seems to be a natural, a perfect criterion, so he settles on growth and development with the Wetzel Grid as the criterion. Now interest and education are wedded, and the topic becomes "The Effect of Interscholastic Football Upon the Growth of Junior High School Boys, as Measured by the Wetzel Grid." Later the difficulty of getting co-operation and of controlling the many variables leads the researcher to add one more delimitation—he will take only his own school into consideration. Finally, the problem has become "The Effect of Interschool Football Upon the Growth and Development of Boys of Lake Junior High School as Measured by the Wetzel Grid." Quite likely the adviser will think of future librarians and of the difficulties with the bookbinder and may suggest a title of "Junior High School Football and Its Effect Upon Growth and Development." The limitations will, however, be given early within the text of the thesis and the problem remains the same.

Personal Criteria. These are interrelated and supplementary in nature. *Interest* in a proposed study is indispensable to its successful completion. Faint curiosity or even the fact that the project seems to be the least distasteful of the lot may ultimately change to a positive interest, but these are weak reasons for starting on the most discriminating and revealing of professional hurdles required for the higher degree. A real interest in his topic will carry the student over obstacles which the more negative reasons might render too formidable or professionally fatal. To have a definite interest in the topic for its own sake, not just as a means of meeting an educational requirement, is good insurance that he will finish the task, and creditably so. This interest should grow as the student becomes more enlightened and informed through his background reading.

Capacity, one's ability to do a given study, obviously is not perfect at the start. There are many unknowns and many areas in which present skill is inadequate. The scholar is driven by his interest and curiosity to delve deeply into related literature. He may even take a course or two, or audit some previous course, now grown hazy. A statistics course, a measurements course, or an additional advanced course specific to the problem can yet be taken, even though not prescribed by the committee. No one has been examined by a psychiatrist for such an act, even if it may seem to be a bit out of the ordinary. The lazy, unscholarly individual to whom the thesis is an inescapable necessity will try to tackle the problem with the least pain possible, will do a minimum of library research, and will take the shortest route. Here is news for that person! Not only is that way not short, it is dangerous. One's reading will reveal techniques successfully employed in similar studies. These techniques may not be at the immediate command of the embryo researcher but he should take down the name of the technique, something of its effectiveness, and any formulas involved. The *how* will come later with study or with other help.

Some scholars have been known to take a major in a new field which was essential to the solution of a particular problem for which there was a great passion. This need is rare and as a rule might be construed as an indication that one should not undertake that problem. One scholar, under the direction of the writer, lacked the necessary advanced statistics course and accordingly bought the needed books and taught himself the required correlational and regressional techniques. This too may not be ordinary, but neither was the scholar. Remember, what happens to the researcher can be of greater worth than his resultant research.

Physical and emotional capacity of the researcher should also be considered. If the researcher is easily frustrated by minute details, great care should be taken in selecting a problem which will not be likely to strain to a breaking point the researcher's physical and emotional capacity.

Feasibility can be made, as well as found, but there are limits, economic and temporal, and these limits are known even to the scholar. He will investigate the costs of equipment, travel, calculations, payment for experimental subjects, and the like. He will compare the probable length of time required for the completion of his proposed study with time reasonably available. His adviser, persons having finished similar studies, and those who might furnish the equipment and calculations, or serve as subjects, will give him indications upon which to estimate the economic costs. These he should balance with his own family and professional plans, with his economic status, and of course with his zeal for the problem. The answer is again only available to the candidate. There seems to be a tendency for a minimal standard in judging what is excessive. On the other hand, problems are frequently more costly than expected, so let the buyer beware. Before giving a negative decision on a vital topic, however, the researcher would do well to explore the possibility of grants, subsidy, or an assistantship. The more definite and vital the proposed study, the more likely the financial aid. Many questionnaire studies, done by mail on an opportunistic group, would be given no consideration, for example.

Availability of data is another consideration closely related to feasibility. Not all available data are relevant; nor will all relevant data be attainable at times. Again, the merits of the study, its sponsorship, and its uniqueness and timeliness will somewhat determine whether some data will be made available or not. Quite frequently, data collected for one purpose (such as physical fitness data collected by the armed services or by college fitness programs during the war) may be too unreliable or incomplete for present needs. Obviously these inadequacies cannot be remedied at a later date nor can fallible data be rendered truthful by statistical manipulation.

Usually data collected for the specific problem under consideration and according to standard specifications are best, but not always. Consideration must be given to the possible demands made upon the time, energy, and privacy of the individuals or institutions. One must judge whether important practices or programs must be set aside temporarily and to what degree. The open end questionnaire with endless queries—frequently overlapping and with answers often available elsewhere—usually results in markedly reduced returns, incomplete responses, or carelessness and inaccuracies.

Personal bias is natural. The researcher should ask himself, "Am I trying to prove that interschool athletics are harmful (or not harmful) to elementary school pupils, or am I trying to assay the truth of the matter? Am I emotionally involved in the direction in which the chips may fall?" This matter of personal bias can unconsciously make differences of tens of percents when real differences in measurement or survey results might be revealed within a few points. Religion, race, state pride or prejudice, or professional zeal may render the researcher incapable of objective measurement, questioning, or interpretation. On the other hand, a cool, conscientious, careful scholar may obtain the necessary drive to a thorough job from just such a motivation. A person can be warned but decisions cannot be made for him. His proposed outline will be quite a revelation of his ability.

Personal returns from the research should be evaluated. The researcher may ask himself, "What will be the personal returns from a task of this magnitude? Will I have a feeling of satisfaction and pride in a job well done? Will there be such a demand for the results by the workers in the profession that it will bring credit to me and my institution? Will this study fill a specific need for a certain field or a group of its workers, so that I may become recognized in this regard? Can the test be published and placed on sale? Does the study lend itself to publication in textbook form? Will my own skills and abilities be so enhanced by this project that I may become confident and better able to pursue my chosen profession or better serve in some of its special functions? Will I be able to teach or to conduct and direct research as a result of this experience? Or, will I be ashamed of my performance and glad to forget it and to have it forgotten? Will I never be led to attempt another study requiring research ability once this chore is over?" Surely more than just meeting a requirement is possible!

If this is not true for the topic under consideration, then perhaps the right topic has not yet been found.

Social Criteria. These are not inseparable from personal criteria. They are, however, more difficult to meet, at least for the novice in research. It is well to indicate these guides as to the social worth or value, for if nothing is ventured, nothing is gained. Also the chances are that in the long run there will be first or introductory research studies which will be a real contribution to the profession. In all truth, most doctoral candidates aspire to make such a contribution through their study.

Fundamental importance of a topic has been said to be determined by "how many people will be influenced and how much." Not everyone can be a Terman whose revision of the Binet-Simon intelligence test gave the first national norms for guidance in educational programs, or a Salk whose vaccine could protect millions. However, tests of strength, fitness, athletic ability (4, 19), motor ability, and sports knowledge (19, 21) have been devised which either came to have national use or which established a criterion and/or procedures for much subsequent research. Growth charts are examples of practical contributions to school health work (23). If a state is served or a real contribution to a city department is made, if new techniques are successfully used or developed for the first time, then a study of fundamental importance has been made. It is impossible to mention all original or stimulative contributions in the areas of health education, physical education, and recreation. Some must be omitted or skipped over lightly, as for example the recent contributions in the field of kinesthesis or the use of statistical designs for causal or factor analysis in sports methods experiments.

Timeliness gives support to efforts which might otherwise be neglected. The present and recent interest in physical fitness, the interest in standards for facilities following World War II, the current debate on interschool athletics below the high school level, sociometric group dynamics, and action research are all timely possibilities.

Uniqueness (13), novelty, or long-sought solutions such as longitudinal growth studies may be desirable. On the other hand, the omnipresent, mailed questionnaire study, so apt to intrigue the nontechnical candidate, is an example of an overdone and frequently poorly executed approach which is unlikely to receive great acclaim.

Research facilitating other research (8) is always acceptable, if carefully done. Co-operative research (2, 12) has given problems to scores of others in our field. Closely related is group research in which several can contribute, such as the setting up of achievement scales (5, 6, 15). In these, hundreds participated. Such studies are opportunities for novitiates to get the feel of careful work for a larger social purpose.

Unification of knowledge can be brought about by synthesis of research in particular areas. Also, in national studies or in cross-sectional studies, the effect of athletics on health in the several school levels or determination of fundamental predictors of sports skills for the myriads of sports are examples of gaps in our professional knowledge which need to be filled.

PROBLEM SUGGESTIONS

There are four frames of reference which can separately or collectively become quite suggestive of potential problems to the beginning research worker seeking a problem to tackle—analysis of the areas or branches of the chosen field, commonly recognized problematical areas, differentiating factors or variables, and possible research methods or techniques.

Field Analyses. There is no assumption that the following field breakdowns are complete or acceptable to the philosophers in these fields. The sole purpose is to give examples and to explore problem suggestions.

Health and safety can be said to consist of:

(a) HEALTH AND SAFETY SERVICE—health examinations, health and safety inspections, immunization, isolation, nutrition work, follow-up and guidance, safety patrols, first aid treatment, and special classes (heart, hearing, sight, crippled, anti-tuberculosis, and the like);

(b) HEALTH INSTRUCTION—personal hygiene, community hygiene, industrial hygiene, social hygiene, mental hygiene, first aid, and family living;

(c) SAFETY INSTRUCTION—home safety, school safety, transportation safety, recreation safety, occupation safety, and driver training;

(d) HEALTHFUL SCHOOL LIVING—hygiene and safety of the school plant, hygiene and safety of instruction, school lunch, healthful and safe routine, and educational guidance.

Physical education may consist of the service program with its administrative and hygienic activities, sports (developmental and recreation), gymnastics, and rhythmics; the adapted-restrictive and remedial program; the coeducational and corecreational programs; the intramural program; and the interschool athletic program.

Recreation may consist of recreation education, school-conducted community recreation, school services to community recreation agencies, outdoor education, community recreation, commercial recreation, and private recreation.

Public health may consist of enactment, enforcement, service, education, and engineering.

This empirical breakdown into some 43 major areas could be further refined until the list would be endless, or reasonably tenfold. Any one researcher would of course be interested essentially in his own field—health, safety, physical education, and/or recreation; hence, the potential and the actual scope would not agree.

Problematical Areas. Within any field or subfield, there are certain common areas of concern. Within these areas, problems of somewhat unique nature are bound to arise. Out of these problems come the need or opportunities for research. Again it is not held that the exemplary areas chosen are a completely logical or acceptable analysis. However, this empirical breakdown has functioned in the past and will serve in this present effort at problem analysis and elicitation.

Philosophical problems deal with purposes, principles, policies, and values and are usually a matter of logic. However, a person's accepted system of philosophy such as idealism, realism, pragmatism, or some blend thereof, will be an essential factor in the results.

Legal or regulative problems deal with mandates, permissions, and the like. One may be concerned with existing or with needed laws or rules. In scope, they may be international, national, state, or local. They may be pol ical or institutional.

Administrative and supervisory problems consist again of principles, policies, procedures, and practices. Their essential pur poses are to set the stage for, or to implement, the programs mentioned above under Field Analyses.

Program and curriculum problems deal with what content, activities, or service should be provided for whom, when, how, in what amount, and in what degree of difficulty.

Methods or *organization problems* deal with *how* the activities or content are to be presented or the participants organized.

Personnel problems deal with the nature of, selection of, guidance of, training of, and evaluation of the success of staff or participants in the various programs.

Facilities problems are concerned with the space, structures, fixtures, equipment, and supplies appropriate to certain fields for certain purposes. Principles, criteria, and standards for planning; and construction, use, and maintenance of these facilities may also be somewhat a concern of philosophy at the over-all level.

Financial problems deal with the source and expenditure of monies and with their accounting in any field.

Relationships deal with the areas of responsibility and authority between fields or programs. Co-operative relationships are the essential phase of this area since legal or regulative and administrative or philosophical relationships are also involved in the original organization of these regulations.

Professional problems deal with the extralegal co-operative efforts to advance the field through codes of ethics, writing, conference, and joint study of professional matters.

It must be admitted that there is some possible overlapping in these ten areas. However, each division contains some unique possibilities for problem study of a distinct nature. Thus, potentially, the 400 to 500 crudely estimated fields and subfields of interest first outlined, when combined with these ten proposed problematical areas, make a theoretical possibility of some 5,000 kinds of problems in all of our fields.

Differentiating Factors or Variables. By means of this frame of reference, the researcher may delimit problems to workable size and receive guidance as to the nature of independent variables to control or possible causal factors to hypothesize and test.

While probably not limitless in number, these factors or variables, contributing to the nature and scope of potential problems, are far more numerous than is indicated here.

Age, sex, height, weight, physique, training, experience, and *intellect* are some of the more personal human variables which may be presumed to be, and in some cases are known to be, causal factors in certain situations.

Race, religion, socio-economic status, and *occupation* are interpersonal and social factors and variables which might well be considered in some problems.

Geography, political unit, auspices or sponsorship, town or institution size are social factors to consider.

Time is a variable in learning, and in growth and development especially. *Levels*—such as beginning, intermediate, or advanced—or grades in school are other conditions to be controlled.

The 18 variables listed are only suggestive of many which reading might reveal to be important. These considerations, with their subclasses and in combination with the 5,000 or more possibilities developed above, assure the problem seeker of a reasonably estimated 100,000 potential sources of problems.

Research Methods. Not all research methods are equally applicable to all fields of interest, all problematical areas, or all variables. However, consideration of the technique or method most adaptable to the problem or to the researcher's abilities will again be a guide in problem selection.

One example of the need for a choice of research method follows. The researcher may be interested in the progression of stunts for the elementary grades. He may glean texts on stunts and tumbling. After applying some experience and logic to similar stunts with differing names, he may tally the frequency of use of stunts for different grades for both sexes. Lacking such grade placement, he may note the order of mention. Out of this maze of differences of opinions by experts and confusions of bases of classification, he may come up with a possible order of progression. He will have used *library technique.*

Not satisfied with this debatable hodgepodge, the researcher may wish to verify or validate the order more fully. With a check-list or questionnaire based upon the obtained sequence, he may by some criteria select some modern "experts" who should be "in the know" as to the relative difficulty of the stunts and request their expert judgment as to the *real* order of difficulty. *If* they are all expert, and *if* they all respond, he will now have a modern "estimate" of the proper order of stunts for the elementary grades. This will have been a *normative survey*, and he will have utilized the *checklist or questionnaire technique.*

If truly scientific, the curriculum researcher might still be dubious and may wish to test "opinion" against actuality. He will accordingly use the authoritatively revised order of difficulty as a best guess as to the probable order of success and safety and will randomly select samples of boys and of girls from each of the appropriate grades. To these, he will present the stunts, recording whether the pupils succeeded upon the first trial, or upon the second opportunity, or whether the stunt was failed. Summing the scores and putting them into order by size, he will now have the true difficulty level of various stunts for *his sample.* Other samples of other populations where greater or less emphasis may have been put on stunts in the physical education program will, in all probability, reveal different difficulty orders for these same stunts. In the present instance, the researcher may be said to have used a *simple experimental procedure.* Thus, we see there are more ways than one to do research.

Methods of potential use in these and similar or dissimilar problems might include philosophical, historical, curricular, survey, measurement, experimental, or many other variant or sub-level techniques. Choice of the technique to use will depend upon the nature of the problem, the researcher's intellectual capacity, the degree of assurance he needs to have in the answer, the account of money or time available, the urgency of the problem, and his zeal for the problem. Whichever method or technique may be employed, the researcher is held for accuracy, objectivity, and thoroughness. Many problems will use several techniques, such as library, measurement, and finally experimental.

Thus, it can be seen that the range of possible problems is probably in the millions. Yet how often does one hear this lament—at least of beginning graduate students—"I just can't find a problem"?

BASIC PRINCIPLES FOR DEVELOPING A PROBLEM OUTLINE

A doctoral candidate, and especially a Master's degree candidate who intends to write a thesis, should give early consideration to his thesis topic. He should carefully explore the possibilities revealed through the application of the personal and social criteria mentioned earlier. Many times problems can be developed and refined in introductory research courses or problems seminars. A problem thoroughly planned is well on the way to completion.

The advice of his chairman and other professors, as well as that of mature graduate students, should be sought, and related literature should be located and carefully gleaned.

One logical and thought-provoking technique for directing the researcher's thinking and planning on a proposed problem is the "horizontal analysis" technique. In essence, this is nothing but a formalization of a simple series of steps: define or state the problem; determine the major subproblems; and for *each* subproblem state precisely *what* one needs to know, *where* one can locate the needed facts or techniques, *how* one will locate them, the proposed *organization and analysis* of the facts, and the resulting type of *conclusions.*

THE PROBLEM OUTLINE

In the problem outline, there should be three major parts: the introductory material, the horizontal analysis, and the bibliography.

The Introductory Material. A few paragraphs or a page or two will suffice to give the motivating reasons for having chosen the proposed problem. The researcher's special aptitude and/or experience in the field, a current demand for a solution of the problem, a gap in related knowledge, or even the availability of adequate and accurate data essential to the solution of the problem may be cited.

A concise and yet revealing statement of the problem should follow immediately. It might be well to state the several major subproblems to further delimit or define the scope of the study.

The *purpose* for the study or the possible use of its results should be outlined. For example, if a performance test is to be developed, the researcher should state for which sex and for what grade level the test is intended. He should also state whether the test will be primarily diagnostic, a measure of achievement, a basis for classification, a research tool, or a combination of these.

The *need* for the test should be stated. For example, in a physical fitness program in which results are expected and must be measured, the *significance* of the study may be cited—national need, contribution to safety, motivational value, or utility under normal teaching conditions.

The *delimitations* or scope of the study should be outlined. For example, in physical fitness research, the outline should include: the number of cases to be used; whether mass, squad, or individually administered; whether it is for males or females; for which grade or grades it will be set up; whether it is for local, state, or national use; whether general fitness or mere skills in a specific activity are to be assayed.

The *limitations* or known weaknesses should be stated. For example, if in a survey the sample is to be opportunistic rather than random; if a questionnaire is to be mailed rather than checked by personal interview; if a limited number are to be obtained; if there are geographical limitations; or if only existing data will be obtainable—these weaknesses should be admitted at the start.

Essential definitions of terms should be formulated, such as: "What is an athlete? When is one a participant? What is meant by physical fitness? Who is to be counted as a freshman or a sophomore? What are considered to be administrative practices? What is an attitude?" And similar terms relevant to the study need to be defined.

Basic assumptions should be given for the problem; there should be an indication of the way in which the data meet the basic assumptions required by the techniques for the analysis of the data.

A *problem* must be *unitary*, that is, deal with one essential goal at a time, and be properly delimited. Furthermore, steps to solve it should be taken one at a time and in strategic order, i.e., each preceding step being indispensable to the attack upon the following step. As a rule, research studies follow this pattern.

A first step in the recognition of this necessary, or at least more effective, order *is the division of the problem into its major subproblems.* Their number will depend upon the nature of the study involved, its inherent complexity, and to some extent the naïveté of the researcher. A true problem can have no less than two such subproblems, will usually have three or four, and conceivably might have more. When there are more than four, the researcher should reconsider to determine whether he may be attempting a multiple study, or perhaps may have made the common error of including as subproblems mere minor steps or "items one needs to know."

An example of choice of subproblems is given in the horizontal analysis of the "Survey of High School Health and Physical Education Programs for Boys in Massachusetts" (p. 64-70). This topic was selected to guide future survey workers in doing a good piece of research, because surveys have usually been poorly done in the past. In this proposed study there are three subproblems. Subproblem I is the selection of and orientation to a score card and related variables. In this step, the researcher is involved in a thorough study of relevant score cards, studies involving their use or criticism, texts in the field, and sources for the evaluation of score cards or of programs. Thus, he is referring to and gleaning authoritative sources. Books and even authorities in the field are the source, and the objective is a sound choice of a score card from among existing score cards.

If there are no such score cards or if the available score cards do not reasonably meet the criteria suggested and the purpose of the study, then a score card would have to be devised, or at least adaptations made, before the proposed problem could be carried out. A basic principle applying here is, "We cannot use what we do not have, nor should we develop that for which we have no use." The result of this first subproblem is an indispensable score card.

In checking upon his analysis, the researcher should be sure that no jury, list of criteria, principles, checklist, data card for hand sorting, categorization for data analysis, definition, or similar treatments are employed in the analysis without first evidencing the manner in which they were developed or attained. Foolish as it may seem, the beginner should also be warned not to go to the trouble of developing such techniques or devices and then fail to utilize them later in the process of developing the problem.

Subproblem II in the analysis is "selection of sample schools and application of the score cards thereto." Here the researcher must have the schools before he can assay them. However, before he can assay the schools accurately, he must have pilot schools willing to allow him to practice on them in order that he may subsequently reliably and objectively evaluate his sample. In this major step the source is *schools* and the objectives are *their scores* and other relevant data. Here, little reading matter would seem to be required other than an atlas for town sizes and a school directory for school sizes and location. It is obvious that from the similar and related studies gleaned in the first subproblem the researcher has discovered pertinent independent variables—the need for thoroughness in scoring and for exactitude in sampling procedures if his findings are to be comparable with those of other workers. The results of this subproblem are the desired values of the school scores on score card items, areas, and total. In some problems, it is desirable to analyze the data from the pilot study to be sure the data can be analyzed in the way planned and that the kind of results desired can be obtained. Changes in the instrument for collecting data may have to be made. Then there is the third and final step.

With the needed scores available, the researcher finally proceeds to Subproblem III in which he "organizes and analyzes the data" so that he may test his hypotheses by the statistical findings and arrive at generalizations and recommendations—the original purpose of the study. In this *final* major step, the categorization and classification of data, so that they may be treated statistically, and the drawing of tenable conclusions or inferences from the findings are a unified and interrelated, but orderly, process of inductive thinking by aid of calculators and statistical formulas. Here the source is statistical treatment of data, and the objectives are tenable and demonstrated conclusions or inferences with their logical recommendations—the purpose for which the problem was conceived in the first place.

Thus, a subproblem is any major indispensable and unique phase of the problem which may be developed to its own conclusion or conclusions by appropriate steps or procedures, which are likewise unique to the subproblem and its purpose. Each subproblem is necessary for the understanding or solution of the next subproblem.

In order to develop each subproblem properly, the researcher must indicate: A. *What* he needs to know or have; B. *Where* he may locate the knowledge or object; C. *How* he will locate the desired fact or thing; D. *How* he will proceed to *organize* and *analyze* or *utilize* that which he has located; and finally E. The *kind of conclusions* expected. The first five of these steps must be explained or described in all necessary detail. What is necessary detail is again determined by the complexity of the step, the naïveté of the researcher, and the consequent requirement of his committee or instructor. The more experienced and well-read the worker, the less the detail (within limits) which might be considered necessary. For example, a trained and experienced statistician might merely say he would test for significance. For one ignorant of the field, the greatest detail would be needed. No mathematical rule can be given but, if in any doubt, it is best to develop the item completely. The last process—the conclusions—can only be given in general as to *type*, or categorically. The conclusions must logically be derived from their preceding "organization and analysis." They obviously cannot be given specifically, or else the items are not something the researcher needs to know. However, he has to have an idea of the end result in order to be able *to recognize it when it is obtained. It* is obvious that the depth of his experience and reading will determine the utility of the outline. If he can proceed from this "blueprint" and unfalteringly take each necessary step to the kind of conclusion expected, there is probably sufficient detail. Allowing his fellow students to look at the outline to ascertain how completely they understand the planned research steps may reveal ambiguities, additional steps, or definitions which may need attention. Let us look at the example of the horizontal analysis again.

Subproblems are first broken down into "What one needs to know (or have)." These steps follow logically from the first item of the first subproblem to the last item of the last subproblem. The mere mention of these steps in their proper sequence does not solve the problem but it does tend to reveal the scope or magnitude and to indicate the strategic order of things to come. These steps are usually in question form to help the researcher in developing the conclusion.

In the survey study outlined on pages 64-70, the first subproblem has been resolved into: "1. What score cards are available? 2. By what criteria shall they be evaluated and chosen? and 3. Which of the available score cards best meets the criteria?" On the right side of the outline under "Conclusions," the matching conclusions are the kinds of accomplished results of the activity in taking the steps under "*What one needs to know.*" The researcher never puts the specific answer in these conclusions, because then there would have been no need for the step. Instead, that fact or object probably would have been a basic assumption in the prefatory material. The conclusions matching the items or steps in the first subproblem are: "1. A list of available score cards; 2. Criteria by which to judge them; and 3. A selected score card." It should be noted that there are no statements of things to do in this final column—only accomplished facts in kind, not specifically stated. Also the Conclusions column reads logically from top to bottom, as should all of the columns if logically done.

Next let us follow the first item one needs to know, across to its conclusions. Under "B. Where will I find it?," it can be seen that the researcher has some 14 specific texts, score cards, or sources. These have all been read, abstracted, and noted on his card index file. This is called gleaning and is mentioned in "C. How will I find it?," and "D. The organization and analysis of the data." Finally, the first conclusion under column E is "an alphabetical list of available score cards with source, cost, and uses." Thus, the researcher has gone logically from a need for score cards to their location and final listing and description in an orderly manner.

For each item the researcher needs to know, there is described in detail exactly what he needs; where it may be obtained; how it will be obtained; how it will be organized, analyzed, or used; and the expected type of conclusion. A close scrutiny of the elaborate analysis will reveal that every step taken from conception to conclusion of the entire problem is explained in logical order. This further reveals that the researcher has read pointedly and critically, and has evolved a "blueprint" or master plan which will guide him continuously, step by step, until the ultimate solution is reached. Thus, a problem that is well planned is already on its way to solution before any further steps are taken.

No one can foresee every possible incident, misadventure, or obstacle that the researcher may meet in his progress. However, the more thorough his preliminary reading and outlining, the less the unknown, unexpected, and unforeseeable will be encountered. Neither can, or should, he list each minutest detail in his outline. For example, just giving a needed statistical formula for a revealed need is adequate, as the bibliography will contain the statistical source to refer to for the explanation and interpretation of statistical processes.

The reader has been guided down column *A* of what one needs to know, and across the first row, "I. What score cards are available?" Each row should be read through from left to right and consecutively from top to bottom in the "Horizontal Analysis," so that the reader may realize how this plan guides the researcher through his study. It would likewise reveal to another researcher or his committee that the problem can be done. As a carpenter follows the architect's plans, the researcher follows his blueprint.

While the researcher is in the process of planning his problem, this horizontal analysis technique would seem to the writer to be indispensable. After the outline is satisfactory to the candidate and is approved by his chairman, the plan might well be written vertically in paragraph form in single page style. To do this, the researcher merely indicates his subproblems, one at a time, and vertically enumerates, one at a time, each item he needs to know; where he will get it; how he will get it; the organization analysis, and utilization; and the type of conclusion or conclusions. When each item has thus been outlined in order for the first subproblem, the next subproblem is stated and the items continued in order as above.

The weakest point in a horizontal analysis will usually be the *Organization and Analysis* column. This column readily reveals the researcher's lack of logic and failure to read in depth and in breadth. He must know and tell *specifically how* he will organize and analyze the facts, items, or data needed to arrive at the concluding statement in the last column.

He should not say "I will tabulate"; "I will develop"; "I will organize." He should say how and under what categories or steps, as the case may be, the data will be processed and analyzed.

Two other problems of a distinctly different type are here outlined as far as the subproblems and the steps for each. These sample problems consist of first, "Steps in Constructing a Performance Test," and second, "Steps in Constructing a Written Test." Only the basic outlines are given. These outlines may serve as the bases for the development of a horizontal analysis.

Steps in Constructing a Performance Test

Subproblem I: How Shall the Test Items Be Determined?

1. Determine purpose of test
2. Analyze ability to be measured

3. Determine criteria for test item selection
4. Select experimental items
5. Select criterion measure(s)

Subproblem II: How Shall the Test Events Be Administered?

6. Construct record form(s) and directions for administering and scoring items and criterion measure(s)
7. Obtain equipment and facilities necessary
8. Conduct a pilot study in item administration
9. Revise items, forms, and directions
10. Select sample of subjects
11. Administer experimental test items and criterion measure(s)

Subproblem III: How Shall the Data Be Treated Statistically?

12. Score test items and criterion measure(s)
13. Analyze test and criterion measure(s)
14. Combine items and obtain multiple correlation with criterion
15. Compute regression equations or sum T-scores
16. Compute norms
17. Make up test manual

Steps in Constructing a Written Test

Subproblem I: How Shall the Test Areas and Items Be Selected and Preliminary Test Set Up?

1. Determine purpose of test
2. Establish curricular validity for areas and items
3. Set up table of specifications
4. Set up criteria for good test items
5. Construct test items
6. Construct preliminary test
7. Set up scoring device or key
8. Set up format and directions for test and scoring device

Subproblem II: How Shall the Preliminary Test Be Administered?

9. Conduct pilot study in administration of test
10. Revise test directions and scoring device in light of pilot study
11. Select sample of testees
12. Administer preliminary test

Subproblem III: How Shall the Final Test Be Determined?

13. Analyze preliminary results—difficulty and discrimination of items
14. Revise test in light of test results
15. Select sample of testees
16. Administer final test results
17. Analyze final test
18. Revise final test in light of analysis
19. Set up norms
20. Make up test manual

For the first subproblem, on obtaining background and determining items, the sources are authorities, books, and dissertations on the subject to provide the knowledge of what one is to work for and with what.

The second subproblem is concerned with administering the test items to subjects to obtain the necessary raw data. The source is the group of subjects taking the test items and the purpose is to obtain necessary raw data.

Finally, the third subproblem is the statistical analysis, interpretation of data, and writing up the research. The source is the raw data treated statistically, and the result is the selected test, its norms, etc.

Roughly, authorities, subjects, and statistics comprise the three subproblems in rather clean-cut delimitations. Next, it is necessary to know for every item (a) the *specific* source or sources for such information or objects; (b) exactly how to obtain such necessities; (c) the *detailed* techniques and procedures for organizing, analyzing, and utilizing these facts or objects after they are obtained; and (d) finally, the *type* of conclusion which might be expected out of this manipulation or use. Each subproblem will be resolved within its own scope and its solution will be necessary to the succeeding subproblem or subproblems. At the conclusion of the final subproblem, the solution to the original problem will have been accomplished.

A helpful exercise for the person interested in research might well be to attempt to get the necessary background from the literature and work out the complete analysis for one of these suggested problems.

BIBLIOGRAPHY

1. AMERICAN ASSOCIATION FOR HEALTH, PHYSICAL EDUCATION, AND RECREATION. *Research Methods Applied to Health, Physical Education, and Recreation.* Washington, D. C.: the Association, a department of the National Education Association, 1952. 535 p.
2. BOOKWALTER, KARL W. "A National Survey of Health and Physical Education for Boys and Girls in High Schools, 1950-51." *American Academy of Physical Education Professional Contributions No. 4.* Washington, D. C.: the American Association for Health, Physical Education, and Recreation, a department of the National Education Association, 1955. p. 1-11.
3. BROWNELL, CLIFFORD LEE. "A Scale for Measuring Anteroposterior Posture of Ninth Grade Boys." *Teachers College Contributions to Education 325.* New York: Bureau of Publications, Teachers College, Columbia University, 1928. 52 p.

4. COZENS, FREDERICK W. "The Measurement of General Athletic Ability of College Men." *Physical Education Series*, Vol. 1, No. 3. Eugene: University of Oregon Press, April 1929. p. 125-94.

5. COZENS, FREDERICK W.; CUBBERLY, HAZEL J.; and NEILSON, N. P. *Achievement Scales in Physical Education Activities for Secondary School Girls and College Women*. New York: A. S. Barnes & Co., 1937. 165 p.

6. COZENS, FREDERICK W.; TRIEB, MARTIN H.; and NEILSON, N. P. *Physical Education Achievement Scales for Boys in Secondary Schools*. New York: A. S. Barnes & Co., 1936. 155 p.

7. CRAWFORD, CLAUDE C. *The Technique of Research in Education*. Los Angeles: University of Southern California, 1928. 320 p.

8. CURETON, THOMAS K. *Standards for Testing Beginning Swimming*. New York: Association Press, 1939. 185 p.

9. DYER, JOANNA T. "Revision of the Backboard Test of Tennis Ability." *Research Quarterly* 9:25-31; March 1938.

10. ESPENSCHADE, ANNA; DABLE, ROBERT; and SCHOENDUBE, ROBERT. "Dynamic Balance in Adolescent Boys." *Research Quarterly* 24:270-75; October 1953.

11. GOOD, CARTER V., and SCATES, DOUGLAS E. *Methods of Research, Educational, Psychological, and Sociological*. New York: Appleton-Century-Crofts, 1954. 920 p. (See also GOOD, CARTER V. *Introduction to Educational Research*. New York: Appleton-Century-Crofts, 1959. 424 p.)

12. LARSON, LEONARD; FIELDS, MOREY R.; and GABRIELSON, MILTON A. *Problems in Health, Physical Education, and Recreation Education*. New York: Prentice-Hall. 1953. 340 p.

13. LEYHE, NAOMI. *Attitudes of the Women Members of the American Association for Health, Physical Education, and Recreation Toward Competition in Sports for Girls and Women*. Doctor's thesis. Bloomington: School of Health, Physical Education, and Recreation, Indiana University, 1955. 270 p.

14. McCLOY, CHARLES H. "The Factor Analysis as a Research Technique." *Research Quarterly* 12:22-23; March 1941.

15. NIELSON, N. P., and COZENS, FREDERICK W. *Achievement Scales in Physical Education Activities for Boys and Girls in Elementary and Junior High Schools*. Sacramento: California State Department of Education, 1934. 177 p.

16. PHILLIPS, MARJORIE, and SUMMERS, DEAN. "Relation of Kinesthetic Perception to Motor Learning." *Research Quarterly* 25:456-69; December 1954.

17. SCOTT, M. GLADYS. "Achievement Examinations for Elementary and Intermediate Tennis Classes." *Research Quarterly* 12:40-49; March 1941.

18. SCOTT, M. GLADYS. "Measurement of Kinesthesis." *Research Quarterly* 26:324-41; October 1955.

19. SCOTT, M. GLADYS, and FRENCH, ESTHER. *Evaluation in Physical Education, Better Teaching Through Testing*. St. Louis: C. V. Mosby Co., 1950 348 p.

20. SMITH, HENRY L. *Educational Research*. Bloomington, Ind.: Educational Publications, 1944.

21. SNELL, CATHERINE. "Physical Education Knowledge Tests Developed by the Department of Physical Education for Women, University of Minnesota." *Research Quarterly* 6:78-94; October 1935.

22. WERKMEISTER, W. H. *The Basis and Structure of Knowledge*. New York: Harper & Brothers, 1948. 451 p.

23. WETZEL, NORMAN C. *The Treatment of Growth Failure in Children*. Cleveland: Newspaper Enterprises Association Service, 1948. 102 p.

24. WHITNEY, FREDERICK L. *The Elements of Research*. Second edition. New York: Prentice-Hall, 1950. 539 p.

25. YOUNG, PAULINE V. *Scientific Social Surveys and Research*. New York: Prentice-Hall, 1949. 621 p.

EXAMPLE OF A HORIZONTAL ANALYSIS OF A PROBLEM

Survey of High School Health and Physical Education Programs for Boys in Massachusetts

	A What I Need To Know	B Where Will I Find It?	C How Will I Find It?	D How Do I Organize and Analyze Data?	E Conclusions
I. Selection of and orientation to a score card and related variables	1. What score cards are available?	1. AAHPER, *Introduction to Research in Health, Physical Education, and Recreation*, "Survey Methods" 1952 Dearborn, *Check-list* Good, Barr, and Scates, *Methodology of Educational Research* LaPorte, *Health and Physical Education Score Card No. II* Neilson, *California Score Card* Neilson and Hall, *Utah Score Card* Ohio State Dept. of Public Instruction, *Ohio State Score Card* Other such state surveys, e.g., Bonsett, DeVoll, Fritz, Hall, Lauder, Owens, White, et al.	1. Gleaning references	1. List the score cards alphabetically with source, pages, and cost and intended use. A card index will include all these data under a subtitle "score card," at least one card to a score card but usually several cards will be needed.	1. An alphabetical list of available score cards with their source, cost, and purported use.
	2. By what criteria shall they be judged?	2. Within the limitations of the study As evidenced in such refer-	2. Reflective thinking Gleaning references	2. List criteria as to indispensable, such as 1. Validation by authorities 2. Use requires only a few hours at each school	2. Criteria gleaned from literature and adjudged indispensable and desirable for score

		ences as shown in B, I, 1 above Time and economy considerations	Reflective thinking	. Capable of numerical and statistical treatment 1. Permits of comparison with other state and national studies List criteria as to desirable, such as 1. Validation by discrimination of items 2. Established reliability in other studies 3. Capable of reliable use in present study 4. Capable of objective use in present study 5. Permits of total, area, and item score analysis and comparison 6. Partial compliance with standards can be credited 7. Items properly weighted 8. Reasonable cost of score cards 9. Convenient format, size, etc. (pocket use)	card.
	3. Which score card best meets indispensable and desirable criteria?	3. Score cards in B, I, 1 and criteria from D, I, 2 above	3. Counting criteria met by each score card as evidenced by literature B, I, 3. Logically weight indispensable and essential criteria and select the score card with the highest score.	3. Chart criteria on left and references alphabetically at top, checking each criterion met by each source Select score card meeting 1. most of the indispensable criteria 2. most of the desirable criteria	3. The selected score card for boys health and physical education programs in high schools
II. Selection of sample schools and application of score card	1. Is the surveyor capable of objectively and reliably applying the score card?	1. From pilot application of score card to 6 or 8 convenient and cooperative schools	1. a. Carefully review all items of the score card and visit at least two high schools with two or more companion surveyors and each apply the score card. One surveyor should if possible be *experienced* in the application of the score card.	1. a. Compare ratings on all items and conduct critique of surveyors, determining reasons for any differences in scoring on any item. Come to mutual agreement as to proper score, revisiting if necessary to recheck. If written *interpretations* of any items are needed, duplicate them in convenient form and in adequate number.	1. a. Objectivity in applying score card

	A What I Need To Know	B Where Will I Find It?	C How Will I Find It?	D How Do I Organize and Analyze Data?	E Conclusions
			b. Surveyor after training for objectivity will visit a new school and score it alone. After two weeks, he will return and re-score this school without reference to his previous scoring.	b. Visit additional schools and apply score cards. Compare each item score and determine degree of discrepancy in scores given. c. Continue to visit one school at a time until 95% agreement is attained; this will be considered an objective scoring. (At all times note reason for discrepancy in scores.) d. The surveyor will revisit one of the schools upon which 95% or better objectivity has been attained and will compare his previous and final scores. Ninety-five per cent agreement will constitute reliable scoring.	b, c, d. Reliability and objectivity in applying the score card
	2. What is the accreditation, school size, location, and town size of each public high school in Massachusetts?	2. a. School directory for State of Massachusetts. b. Town sizes in Rand McNally atlas for Massachusetts.	2. a. Glean name, enrollment, and location of all public high schools in current Massachusetts School Directory. Also get principals' names for each. Write on small cards for hand sorting. b. Determine town size (population) of all towns in which schools are located in current Rand McNally Atlas for Massachusetts. Write on respective cards (above).	2. With a convenient scale of town sizes of 8-9 intervals (see other studies in [illegible]) on the top and a convenient scale of school sizes of 6-7 intervals on the side, construct a scattergram and distribute all cards for [illegible] into their proper cells in this scattergram. Shuffle well all cards in each cell and then number cards in each cell from 0 to (N-1) in series. Determine $\frac{100 \text{ (number of schools in sample)}}{N \text{ (total number of all schools)}}$	2. a, b. A list of school their enrollments, and their town sizes in Massachusetts ready for a random stratified sampling

	3. What shall be the random stratified sample of 100 high schools to be visited?	3. From the scattergram in II, XX, 2	3. By selection from a table of random numbers.	3. Multiply the number of schools in each cell by the obtained proportion. These products are the number of schools to be chosen. If n in any cell is too few to permit of a school being chosen by this proportion, combine sufficient adjacent cells *of the same town size to permit at least* one school to be chosen. From a table of random numbers *for each cell* select all numbers from zero to the proportionate number allotted as they occur in the random list. These randomly chosen numbers will indicate the schools which will constitute the random stratified sample. One might wish to choose one supernumerary school in each cell, to be visited, only if some school would not permit surveyor to visit and score.	3. A random stratified sample of all public high schools in Massachusetts numbering approximately 100
	4. What shall be the itinerary for visiting all these schools?	4. From a current map of the State of Massachusetts.	4. By pinpointing all of the sample schools on the state map.	4. Plan a series of trips to and from 2 to 4 schools each. Give consideration to size of school, size of town, distance from base and known conflicts in dates due to scheduled games, vacations, and others. Plan to arrive at each school in advance of scheduled time to learn name of principal and instructor. Leeway must be allowed for schools missed by time conflicts to be picked up later.	4. A planned itinerary most direct and economical of travel, meals, overnight lodging, etc.
	5. What is the status of each school?	5. From the principal and instructor at the schools to be visited.	5. By conference and recording on the score card all facts revealed	5. Checking the appropriate score or fact on the score card as given in the interview and checked by observation of the facility, supplies, etc. A convenient order will be determined for asking the questions as one gains in experience. A copy might be made of the results before leaving and the copy given to the school. This also provides a check that all items have been scored or recorded.	5. The completed score card and additional facts for each school in the sample.
III. Organization, analysis, and interpretation of data	1. A form on which the raw scores from the score cards may be inserted on one convenient card for sorting.	1. From the total score, areas, items, and variables and descriptive information on the completed score cards.	1. By gleaning the score card in II, XX, 5 and by arranging spaces at top of the data card that are of general nature	1. Exemplary data card School ——— Town ——— Accredited Code Code Yes—No— Enrollment ——— Population ——— Size Code Size Code Areas	1. A single card on which the raw data from the score cards may be easily read and sorted.

A What I Need To Know	B Where Will I Find It?	C How Will I Find It?	D How Do I Organize and Analyze Data?	E Conclusions
		and by eliminating discussional materials and substituting coded numbers and spaces for checking or recording the information on the cards, such as:	Items I II III IV V VI VII VIII IX X 1 2 3 4 5 6 7 8 9 10 School Score Totals	
2. Total scores must be analyzed a. What is the rank order of schools by total score?	2. From the data card in B, III, 1	2. By gleaning the data cards and sorting by hand or by IBM if Hollerith cards and machines are available.	2. a. Arrange the cards from high to low and make a bar graph with bar lengths proportional to scores.	a. Rank order bar graph of schools and interpretations.
b. What is the mean, range, and standard deviation of total scores?			b. $\frac{\Sigma X}{N}$, high score — low score, $\sqrt{\frac{\Sigma X^2}{N} - \frac{(\Sigma X)^2}{N}}$	b. The mean range and standard deviations, and interpretations.
c. What is the relationship of accreditation?			c. Find the mean of accredited schools and nonaccredited schools $M_1 - M_2$. Find the errors of the means $\sqrt{(\sigma_{M_1})^2 + (\sigma_{M_2})^2}$ Use "F" test and "t" test for significance.	c. The significance of differences between accredited and non-accredited schools.
d. What is the relationship of town size?			d. Do the same for categories of school size.	d-e. The significance of differences between each school size category, town size.
e. What is the relationship of town size?			e. Do the same for categories of town size.	
f. What are the state norms for total score?			f. Construct a cumulative frequency or ogive curve for determining percentile ranks and scores.	f. Norm (ogive) for comparing status of other schools with sample.

3. Area scores must be analyzed a. Rank order of schools (high to low) by each area score.	3. From area scores on data card	3. Same as above C, XIV, 3.	3. a. Rank schools from high to low in each area by area score. Make 10 bar graphs, one for each area, using school code number to designate schools. Place highest score on top, lowest on bottom, etc. Total length of each bar equal to 30 points. Shade in bar areas proportional to area achievement of total score.	3. a. Rank of each school on each area, by bar graphs.
b. Mean, range, and standard deviation of scores for each area.			b. $\frac{\Sigma X}{N}$, high — low, and $\sqrt{\frac{\Sigma X^2}{N} - \frac{(\Sigma X)^2}{N}}$ for mean, range, and standard deviation for each area. Make a bar graph of the ten areas, ranging the areas by mean area score from left to right, each bar showing mean, σ and range of sample scores.	b. Mean, range, and standard deviation of each area attainment.
c. Significance of difference of means for each area by accreditation, school size, and by town size.			c. Apply the "F" test for homogeneity of variance for each area for groups under accreditation, school size, and town size. Where applicable, apply "t" test to difference between means of the groups, one score card area at a time and one variable at a time.	c. Significance of difference between means for each area by accreditation, school size, and by town size.
d. Degree of relationship of areas to each other and to the total score.			d. Find the zero-order correlations between each area and every other area and each area and total score. $r = \frac{\frac{\Sigma XY}{N} - M_x M_y}{\sigma_x \sigma_y}$ Determine two or three highest r's and two or three lowest r's with total scores and also for area intercorrelations.	d. Important zero-order correlations between areas and areas and total score.
e. Massachusetts state norms for each area of score card (ogives)			e. Set up cumulative frequency curves (ogive) for each area with possible raw scores at bottom and percentile score at right side. Draw vertical lines at median and horizontal lines to the 80th percentiles.	e. Ogive curve for each area.

A What I Need To Know	B Where Will I Find It?	C How Will I Find It?	D How Do I Organize and Analyze Data?	E Conclusions
4. Item scores must be analyzed a. Rank order of item score attainment.	4. From the item scores on the data card in B, XXI, 1.	4. Same as above, C, XXI, 2.	4. a. $\frac{\Sigma X}{N}$ for each item score and rank items, as IX₁, V₄, etc., according to mean item score, giving actual mean beside each item, highest first and lowest last. (Group items in table by single spacing in groups of 5 and double spacing between.) Each column can have thus 25 item means and there will be four columns. Top 25, above median 25, below median 25, and bottom 25 item means.	4. a. Rank order of item means in columns of 25 items each. (Item numbered as in score card.)
b. Degree of discrimination of items of the score card on area and on total score.			b. 1. Select top 27% of schools when ranked on total scores (successes). Select also bottom 27% of schools (failures). Consider item scores of 0 or 1 as item failures, and scores of 2 or 3 as item successes. 2. Calculate Votaw position for each of the 100 items and plot on the Votaw curve plotted by formula $r_t = \frac{2p_t + h^2 \pm \sqrt{(4h^2 + 4h^4)(p_t - p_t^2) + h^4}}{2(1 + h^2)}$	b. Degree of item discrimination for all items and recreation areas.
c. Influence of accreditation, school size, and town size on score attainment.			c. Influence of variables on items 1. Make a table as in D, III, 4a for accredited and nonaccredited items in double columns so positions and values of means can be compared. 2. For each area make line graphs for item means in each school size category. 3. Do the same for town size categories.	c. Influence on item means of accreditation, school size, and town size.

CHAPTER 4

Populations and Samples

FRANCES Z. CUMBEE
CHESTER W. HARRIS, CONSULTANT

TAKING A SAMPLE FROM SOME WHOLE (AGGREGATE, POPULATION) is not a new concept or practice. It has probably been practiced in one form or another since life existed. The cook in her kitchen tastes (*samples*) by a spoonful her well-stirred pot of beef stew (*population*). A potential purchaser may ask for a *sample* from a bolt (*population*) of cloth. A student in college may ask several of her friends (*sample*) out of a list of all her friends (*population*) their opinions on a certain matter. A telephone company may check a certain number (*sample*) of its employees (*population*) relative to the efficiency of their service. Or, the telephone company might ask what proportion of their telephone poles are in need of repair in 1,000 miles of telephone poles (*population*). By some specified plan the officials of this company might select a specified number of poles (*sample*) to get an estimate of this proportion.

A state or national conservation department might select certain plots or areas (*sample*) containing pine trees to determine the proportion of all pine trees in a particular forest (*population*) containing a disease. The laboratory technician takes a drop of blood (*sample*) to ascertain the composition of all the blood (*population*) in a particular individual. A physical education teacher will *sample* three baseball distance throws for a student rather than measuring all throws (*population*) that the student could make in a certain period of time. And so, one could proceed *ad infinitum* to enumerate examples of some form of sampling. The point to be made, however, is that sampling simply means selecting some relatively small number of items, individuals, objects, or the like in order that something may be found out about the population.

The reader can readily recognize the hopelessly formidable task of examining completely the populations listed above. By the time every unit of some populations could be examined, studied, and summarized, the information would probably be out of date and useless. The number of such investigations that could be undertaken would be limited not only because of the time involved but also because of the cost of undertaking such a task. Some short-cuts are needed if information is to be made available in time to be of value and at a cost that can be borne. In all branches of science, some economy is urgently needed in assimilating, interpreting, and understanding the results, if knowledge is to be advanced at some rate other than the pace of a snail. Reducing the number of observations that need to be made allows more time and effort to be devoted to securing more information from one investigation.

The use of a sample to study phenomena has been, and is yet in some circles, viewed with some suspicion and skepticism. Perhaps such skepticism is caused, on the one hand, by being led to wrong conclusions through the use of carelessly applied sampling methods. On the other hand, skepticism may in part be a hesitancy to learn and adopt newer methods. But, wrong conclusions may result also from a study of the population if care is not exercised in processing the data. Many sources of error are possible in basic data. If these sources of error are not carefully controlled, the results from studying all units may be grossly misleading. The reader probably is painfully aware of the errors that can be made in adding a column of figures. Handling large masses of data increases the danger of errors in tabulation and calculation. Even calculators have to have the correct keys punched. Fatigue from handling large masses of data may result in careless punching of keys. These sources of error can more effectively be controlled when smaller numbers are used. In summary, then, there seem to be at least four major advantages in using proper sampling methods: less expense; more speed in processing data and presenting results; feasibility of securing more information from one investigation; and more accuracy, with known precision which may be specified in advance and calculated from the sample itself.

SAMPLING THEORY AND PROCEDURES

Sampling theory and procedures have made great advances in the last few years. One may note that the selected references at the end of this chapter are dated since 1949. In no way is the brevity of the bibliography to be taken as an indication of the lack of vigor on the part of mathematical statisticians in attacking the problems involved in sampling, since the volume of material written on the subject also has increased rapidly in the last decade. Many of the rapid advances have come as the statistician worked on problems arising in large scale surveys. The purpose of the sampling theory so developed has been to make sampling more efficient. In other words, the problem has been one of finding methods of selecting samples and methods of estimating population values that are precise and at a minimum cost.

The joint procedure of selection and estimation are spoken of as the *sampling design*. Adequate planning of the sampling design is fundamental not only in surveys but also in experimental studies. Perhaps it is axiomatic to say that investigations can be no better than the sampling design.

In spite of the rapid advances that have been made in recent years, Stephan (13) points out that much remains to be done. It seems reasonable to suspect that many more refinements will be developed in the future. Certainly one should be on the alert for refinements in present practices and for the development of new theory. To the student with limited mathematical training, the elegance with which the mathematical statisticians present theory is indeed awe inspiring, if not frightening. To the prospective researcher this should not be grounds for utter despair. Fortunately, some of the newer statistics texts (such as reference 14, for example) have presented material in such a way that one is introduced without extensive mathematical treatment to sampling theory, and some of the concepts that help in understanding the theory. Moreover, rather detailed texts (1, 6, 7, 15) are available which help with the application of sampling techniques. At the outset of any investigation, the beginning researcher should consult an authority who has had experience in planning sampling designs. Many decisions must be made prior to the collection of data if the investigation is to be more than an exercise in juggling figures or practice in computational procedures. The point cannot be overemphasized that utmost care is essential in the planning stages of an investigation. Cochran (2) has strongly stressed this point in a good discussion on *salvaging* data. A *salvaging* process would not be necessary if the proper precautions were taken prior to selecting a sample.

On the next few pages will be summarized some of the procedures of sampling. In no way should this presentation be used to supplant the discussion in the texts referred to above. Rather, it is hoped that this presentation will introduce some of the principles of sampling and their applications to investigations in health, physical education, and recreation. It is hoped that this presentation will aid the reader in understanding some of the more technical presentations as well as point out precautions that should be taken. The student may wish to consult references 2 and 7 (Vol. 1, chs. 1-3), 8, 9, and 12 for additional overviews of sampling procedures and their application. Terminology used in sampling will be clarified as the terms arise.

Perhaps one of the first terms which needs clarification is the term *population* or *universe* as it is referred to in several texts. As may be seen by the examples in the first paragraph of this chapter, the concept of "population" as used there is different from the popular notion that "population" refers only to individuals. As used in statistics, *population* refers to the entire group (all units) having some common characteristic. These *units* may be objects, materials, individuals, attributes, deeds, organisms, animals, et cetera. The number of units in a population may be small or large, finite or infinite. In practice, populations from which samples are drawn for study are in actuality all *finite*. It would be impossible to select a sample of pine trees from an *infinite population* of pine trees. An infinite population of pine trees actually does not exist. Such a population exists only as a concept which envisages all pine trees that have existed in the past, all existing pine trees now, and all pine trees that will exist in the future. The reason for which the sample is selected will determine whether the concept of an infinite population is assumed, however, even though in practice one can select a sample only from a finite population. To understand this statement more thoroughly, it will be necessary to examine the two theories of sampling.

Theories of sampling deal with two problems: *enumeration* and *analysis*. The enumerative problem deals with the composition of the population as it is. There is no concern for why the population is this way. The enumerative problem, then, may well be based on a *finite* population. On the other hand, the analytic problem deals with the *cause* system of the population. There is concern for *why* the population is the way it is. An interest in how the population got to be as it is, in order that future populations can be predicted and regulated, must be concerned with the population in the past, the present, and the future. Consequently, *analysis* theory assumes an infinite population. It is also easy to see that even a complete count of the present population would be only a sample of the result of the cause system. Older sampling theories were based on the concept of an infinite population. In general, enumeration theory presented in recent years has been based on a finite population. Care should be exercised, therefore, that formulas used for estimation are those appropriate to the sampling theory applied. For a more thorough discussion of the distinction between enumerative and analytic studies, the reader should see Deming (6: 247-61).

Yet another comment should be made about the sample and its relation to the population. Several of the examples, i.e., the spoonful of beef stew, the sample from the bolt of cloth, and the drop of blood, are only one unit or one "chunk" from the total. It seems reasonable to assume that the units which make up the total population in these instances are fairly uniform. If the assumption of uniformity is correct, the problem of securing a "good" sample does not seem to exist. But, the other examples, i.e., the opinion of friends, the efficiency of employees in a telephone company, the state of repair of telephone poles, the extent of disease in pine trees, and the distance a baseball may be thrown are known to vary considerably. Variability is marked among units in many populations. It is in these instances when considerations of a "good" sample are of extreme importance.

SECURING A GOOD SAMPLE

What then, are some considerations in securing a "good" sample in an investigation in health, physical education, and recreation? Obviously, one of the first questions to be asked is, What is to be investigated? About what is information needed? Is the information to be used only for the sample used in the investigation, or is the information to be used to *infer* some general result for a larger group? If the question is one of knowing what is the average distance a baseball can be thrown by a tenth grader in High School X or the average speed with which a sixth grader can run a 50-yard dash in Elementary School Y, then the question of sampling is not too important. Naturally the answer is to measure, by the best measurement techniques known, all members of the tenth grade in High School X and all members of the sixth grade in Elementary School Y on the baseball throw and the 50-yard dash respectively. Taking measurements on all members in the two classes, particularly if the numbers do not run to 100 or more, and computing the mean ($\bar{X}$) for each group is the thing to do. But herein lie some real mistakes in experimentation and statistical inferences that logic cannot condone. All students in the tenth grade of High School X and all students in the sixth grade of Elementary School Y *are not* to be interpreted as typical (representative) of all tenth and sixth graders, respectively, in the United States or in the world. In the question posed, the population *is* the tenth and sixth grade classes respectively in High School X and Elementary School Y. A *census* of the population has been taken and no further generalization can be made.

Researchers and users of research should adhere strictly to the following rule: *the results of data from any sample may not be generalized outside the population from which the sample is taken.*

If, on the other hand, the physical education teacher asks about the average distance a tenth grader could throw a baseball and the average time it took a sixth grader to run the 50-yard dash in some particular state, or even in the United States, then the question of sampling becomes important. The problem of measuring all tenth and sixth graders in the state or the country is impractical. By the time such measurements could be made and tabulated, and appropriate calculations made, the chances are that the tenth graders would be twelfth graders and the sixth graders would be eighth graders. The practical thing to do would be to select a sample from the state or nation and use the sample to *estimate* the average for these populations.

Many things should be taken into consideration before making a selection of the sample suggested above. Any condition that might influence the performance of tenth and sixth graders, respectively, on the throw and dash should be considered. Such things as the size of the schools, the presence or absence of a physical education teacher, the socio-economic background of students, and others should probably be reviewed. These conditions are not examined, however, to serve as a basis for selecting a sample that is a "little replica" of the population. It is extremely unlikely that a sample will be a "little replica" of the population. If such were the case, the mean of the sample would be exactly the mean of the population and the problem would be solved. Rather, these conditions are examined to help in designing the study so as to reduce the error from sampling.

Suppose such a sample is selected. This situation should be observed more closely. The mean, symbolized by $\overline{X}$,[1] which is calculated from this sample, will be known. But the concern here is to get at the true value for the population which is unknown. How is the researcher to bridge this uncertainty? How may he have any assurance that the sample mean is even near the mean of the population? The precautions taken to see that all groups in the population are considered is not enough to complete the answer. He still cannot be certain that the mean of this sample is the true mean of the population. Is there any way to span this gap?

Probability theory has taught much in this area. The *simple random sample* drawn in this example is a "chance" sample. If another random (chance) sample of the same size (number) from the same population were drawn, it very likely would have a different mean ($\overline{X}$) and a different standard deviation (s).[2] If sample after sample were selected in this manner until all such samples possible were taken,[3] one could make a frequency distribution for the mean of each sample. Such a distribution is a

[1]The student should be cautioned that symbols used in formulas vary considerably from text to text. All precautions should be taken to determine what system of symbols is used by the author. In this chapter, symbols are in parentheses immediately following the term. Tables also have a column for symbols and formulas.

[2]The student is no doubt familiar with the term *standard deviation*. It is used to describe the variation of the measures in a *sample*. A particular point is made of the terminology here because it is to be contrasted with a similar statistic applied to a different type distribution.

[3]It is possible to know how many such samples are possible. This is to be discussed later.

sampling distribution of means; it is a distribution of a *statistic.*[4] From the sampling distribution one could determine how the sample means have varied.

Again, the reader is reminded that these are "chance" means; they are the result of drawing sample after sample of specified number from a population by a random method in such a way that the "chance" or probability of an individual being included in the sample is known. In simple random sampling, each possible combination of specified size has an equal chance of being included. But, rather than denoting the variability of the means in the sampling distribution by the term "standard deviation" (s), the variability is denoted by the term *standard error* (est. σ_M)—i.e., the means have distributed themselves by so much sampling error.

Selecting sample after sample in the manner described above in order that the unknown true value for the population may be determined again is not practical. Is there no solution? Fortunately there is. If the sample has been drawn by a random method so that each combination of randomly selected units has an equal chance of being selected, then some span across this gap between the known and the unknown is possible. Statisticians have shown how the *sampling distribution* (probability distribution) can be constructed. They have presented formulas by which standard errors of the sampling distribution may be estimated. Thus it is possible to usefully approximate what means are possible as well as the chances of these means occurring. This estimate of the standard error (est. σ_M) of the mean may be computed from data for the sample itself, provided the sample is a probability sample.

In practice it is known and can be demonstrated that the mean tends to have a sampling distribution that is normal.[5] If the sampling distribution is normal, then statements may be made about the population mean with a specified probability of being correct. In other words, the researcher knows mathematically what to expect from the probability distribution. A *confidence*

[4]A *statistic* is a summary of some group character of the sample. Group characters from samples are means, standard deviations, etc.

[5]How close the sampling distribution of the means comes to the normal distribution depends upon: (a) the distribution of the trait in the population; (b) the size of the sample; and (c) the design of the sample. If there is some reason to suspect that the sampling distribution is not normal, then another distribution should be sought. Most of the newer statistics books discuss the sampling distributions most commonly found in practice, indicate the conditions under which they may be anticipated, and give the appropriate probability tables to use.

interval with upper and lower limits is calculated. Then, from the probability distribution, he may make the statement that so many times out of 100 the true mean will be between these two values. The confidence interval, based on statistical theory, is the only span across the gap from the known sample mean to the unknown population mean. These confidence intervals are determined using the proper estimates of the standard error. An estimation of the standard error may be calculated validly if a probability sample has been selected. Thus, the deviation of the sample mean from the true mean of the population *may be known with specified degrees of confidence only if a probability sample has been used.*

METHODS OF SELECTING A SAMPLE

Quite logically, then, the next question seems to be: How may a probability sample be secured? The different probability sampling designs are to be discussed in a later section. This section will be devoted to methods of selecting a sample.

A requirement is that every distinct sample of a given size has a known chance of being drawn. Consider this question: How many different samples of size 2 can be drawn from a population of 10? In order to answer this question, it is necessary to specify how each sample will be drawn. Let us agree that we will draw the first member of the sample at *random* (a term which will be discussed below) from the 10, and that we will draw the second member of the sample at random from the remaining 9. This is random sampling without replacement. If we do this, we can determine the number of distinct samples and we will find that each such sample has an equal chance of being drawn.

Students have probably seen the formula

$$\binom{N}{n} = {}_{N}C_{n} = \frac{N!}{n!\,(N-n)!}$$

for the combination of N things taken n at a time. But, as a review here, the formula will give the number of such samples that can be drawn. The reader may wish to review the use of this formula in Cochran (1:11-12) and Walker and Lev (14:18-19). The Walker and Lev reference also explains the use of the factorial symbol, "!". In this example, then: $\frac{10!}{2!\,(10-2)!} =$

$$\frac{10 \times 9}{(2 \times 1)} \frac{\cancel{8} \times \cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{\cancel{8} \times \cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}, \text{ or } \frac{10 \times 9}{2 \times 1} = \frac{90}{2}, \text{ or } 45$$

such samples that can be drawn. The person who wants to make an application of some of the sampling techniques does not have to work out a combinatorial formula such as the one listed here. It is given only as an illustration.

In the simple random sample design, a consequence is that every individual in the population has the same chance to be selected in the sample. As a simple example, suppose that the number of individuals in a population is 10 ($N = 10$). Suppose further that a sample of size 2 ($n = 2$) is to be drawn from the population. If the number of units in the sample (n) is divided by the number of units in the population (N), then the probability of each unit being drawn into the sample is known. In the example, $2/10 = .20$, or an individual has a 20 percent chance of being included in the sample. The formula n/N, then, is actually the *sampling ratio* or the *sampling fraction*. This ratio gives the probability of any individual being included in a simple random sample of specified size.

How does one select or draw a single random sample? The ten names of the individuals in the population could be put on ten tags of equal size and the same surface. These could then be placed in a bowl and well shaken. If somehow one of the tags sticks to another tag, then an equal chance for the ten tags is no longer present. But if no such condition occurred, a blindfolded person could select two tags from the bowl. These two tags would be the sample. If, after the first tag is drawn in this manner, it is returned to the bowl, then we have *sampling with replacement*. Such a procedure would in essence create an *infinite* population. The probability of each tag being selected would then be the same at every draw. However, if the tag is not returned, then we have sampling without replacement, and the remaining nine tags do not have the same chance of entering the given sample as did the first one drawn. At the first draw, the probability of any tag being drawn was 2/10 or 1/5. But, with a sample of size 2, there remains only one draw. Hence, the probability of each tag remaining in the bowl is 1/9. In practice, sampling with replacement is rarely practiced.

Actually, sampling from a bowl seems to be needlessly laborious. Perhaps the easiest way is by use of a table of random numbers. The first prerequisite is to prepare a list (called a *frame*) of all units in the population. This is quite often done by alphabetical order when the population is made up of individuals. A clear description of the use of tables of random numbers may be found in Walker and Lev (14: 126-27). The description will not be repeated here. The reader may wish to review this reference before going on to other sections of this chapter. Certainly a review of this description, or a review of the discussion of the use of tables of random numbers in other statistics books, should be made before trying to select a random sample.

How is one to know that a sample of size 2 in the illustration is large enough to give us the precision desired? Much work has been done on the problem of sample size in relation to some specified cost or some specified precision in sample surveys. Formulas developed depend upon some previous knowledge of the population. These formulas are discussed in several places (1, 4, 6, 7, 15) for the estimation of sample size prior to sampling on the basis of some expected outcome of the survey. These formulas will not be duplicated here, since it is believed that the space might be devoted more profitably to a consideration of the more common sampling designs.[6]

Since the readers of this book are probably concerned with the size of the sample in experiments and hence the necessity of using the appropriate *small* or *large* sample techniques, perhaps some "rule-of-thumb" answer should be given. Authorities rather generally agree that if a sample has 30 or 40 units, then for all practical purposes the sample may be considered a *large* sample. One should be cautioned, however, that the distribution of the trait in the population should be considered in determining sample size. If the trait is highly skewed in the population, a larger sample is needed.

Perhaps it might be more profitable, after the rule of thumb given above, to consider the size of sample in its relation to increasing the precision of the estimated standard error (est. σ_M), the bridge between the known sample value and the unknown population value. Decreasing the estimated standard error increases the precision of the estimate. Kish (9:175-239) has indicated that the only way to decrease the est. σ_M is to increase something. One way to increase the precision is to take a larger sample.

[6]As may be recalled, selection and estimation are referred to as the sampling design.

This, however, is not the only way to reduce the estimated standard error. The different sampling designs have been devised to help with the problem of more precise estimates. The next section will be devoted to an illustration of some of these sampling designs. The illustrations use real data, taken from a defined population. Since it is felt that the reader will gain much insight by actually following through the calculations in these illustrations, appropriate formulas for estimation are given as well as enough summary data.

SAMPLING DESIGNS

Three of the more common sampling designs are to be illustrated below using data[7] from a population for which the parameters[8] are known. The examples to be given are illustrations of survey sampling. Since great strides in sampling techniques have been made in survey sampling, the developments will be presented in this context. It is hoped that having a complete count—a *census* —against which to compare the precision of the various designs will be of benefit to the reader.

All children, boys and girls, in special classes for the mentally retarded in Madison and Milwaukee were given several motor skills tests. The age, height, weight, and IQ of each of these children were recorded. In order that the discussion may be kept brief, only the population of girls and only certain of the measures will be used. The population, then, is of the 103 girls ($N = 103$), chronological ages 8 through 14 years, in 13 schools in Madison and Milwaukee. It is a *finite* population.

PROBABILITY SAMPLES

Simple Random Sampling. Two samples, one of size 10 ($n = 10$) and one of size 40 ($n = 40$), will be selected to illustrate some

[7]The author is indebted to Professors Robert J. Francis and G. Lawrence Rarick for use of these data. Financial support for the project, "Motor Characteristics of the Mentally Retarded," was provided by funds from the U. S. Department of Health, Education, and Welfare, Contract No. 484-2259.

[8]*Parameters* are referred to by statisticians as population values. In this instance, the summary of the population values such as the mean and standard deviation are parameters. This should not be confused with the summary values for a sample, referred to earlier as a *statistic*.

Population values are

	Mean (μ)	Standard Deviation (σ)
Dash	5.90	1.77
Weight	84.78	26.01
IQ	67.31	7.44

points to be made about this sampling design. Suppose that the interest is one of knowing the average weight, the average IQ, or the average speed with which these girls can run 30 yards after a 5–yard start. Prior to making the selection, all 103 girls in the population were arranged in a list (*frame*) by chronological age given in months. Each *sampling unit* (person) was assigned a number in order from one through 103. It was decided that the table of random numbers in Walker and Lev (14:484-85) would be used, entering block one, line one, and reading downward until 10 units were chosen. Since 103 is a three–digit number, the first three columns in block one were used. The numbers drawn into this sample, without replacement, were 94, 103, 71, 23, 10, 70, 24, 7, 53, 5. A summary of the raw scores on the run, weight, and IQ may be found in Table 1.

TABLE 1.—SUMMARY OF SCORES FOR SAMPLE OF SIZE 10

Statistic	Symbols	30-yd. Dash	Weight	IQ
Sum of scores	ΣX	60.70	764	671
Sum of squared scores	ΣX^2	392.81	61,584	45,291
Mean of scores	$\overline{X}$	6.07	76.40	67.10
Variance of scores	s^2	2.71	357.16	29.66
Standard deviation of scores	s	1.65	18.90	5.45

Means for each of the three tests were calculated by summing each of the ten scores and dividing by 10 ($\overline{X} = \Sigma X/n$.) These sample means are estimations of the averages for the population on the dash, weight, and IQ.[9] How may the gap between the known sample means and the unknown population means be bridged? In Table 1 will be found the variance (s^2)[10] and the standard deviation (s) of each of the three measures. Statisticians have shown how the data from a sample may be used to estimate the variance of a *sampling distribution* of all means that would occur if all possible simple random samples of the same size were drawn. Once the estimated variance (est. $\sigma_M{}^2$)is determined, the re-

[9]If the reader follows through with the calculations, some slight difference in answers may be obtained because of a difference in rounding of numbers.

[10]The variance for each of the items was calculated using raw scores with $n - 1$ in the divisor. Formulas for calculating variance (the variability of the characteristic under consideration in the sample) and standard deviation by either raw scores or deviation scores may be found in most statistics texts. After the variance (s^2) is known, one has only to take its square root to determine the standard deviation (s) or $s = \sqrt{s^2}$.

searcher has only to take the square root of the variance to get the estimated standard error (est. σ_M) of the sampling distribution. As indicated earlier, the estimated standard error of the mean from the sampling distribution is the bridge from the known to the unknown.

In general, the formula for estimating the variance to be expected in the sampling distribution is: s^2/n.[11] In other words, the sample variance is divided by the size of the sample. (Since s^2 is calculated by the use of $n-1$ in the divisor, this formula does not have $n-1$ as the divisor for the estimated variance of the mean as may be seen in some texts.) However, the sample in this illustration has been drawn from a finite population *without replacement.* If the "sample," in actuality, had taken all units in the population, then the sampling fraction would be one ($n/N = 103/103$). But only a proportion of that population (10/103) was taken. If the sampling fraction is subtracted from one, then a correction is made for the proportion of the population not included in the sample. The estimated variance must be multiplied by the *finite population correction,* $1-n/N$,[12] in order that a correction be made for not including all units in the population. The formula with the factor for the finite population correction is $(1 - \text{fpc})s^2/n$.

In the example used here, $n/N = 10/103$. One$-$fpc$=1-10/103$. The correction for this example, then, is $(1-10/103) = 93/103$. This fraction is equal to .902. From Table 1 the appropriate values may be substituted in the formula and the estimated variance of the sampling distribution of means calculated. The $(1-\text{fpc})s^2/n$ for the dash is $.902 \times 2.71/10 = .2441$; for weight is $.902 \times 357.16/10 = 32.22$; and for IQ is $.902 \times 29.66/10 = 2.68$. It is an easy step now to calculate the standard error of the mean, since one has only to take the square root of these values. (This quantity is the estimate of the standard error

[11]Estimation of the standard error of a mean is given by the formula, $s/\sqrt{n} = \sqrt{s^2/n}$. Thus, the estimation may be done from the sample standard deviation. It is readily seen that the quantity under the radical is s^2/n, the formula for the estimated variance.

[12]The finite population correction $(1 - \text{fpc})$ may be seen in some texts as written here; in others the factor may appear as $(1 - f)$; yet in others it may be combined with the "n" in the quantity s^2/n and appears as $N - n/Nn\ s^2$. In many studies the fpc may be ignored if the size of the sample is small in relation to the size of the population. Cochran (1) suggests that fpc may be ignored if the n/N factor is no greater than 5 percent. The correction will be used in the illustrations in this chapter.

of the mean. It is the error one expects from sampling.) The standard errors are: the dash, $\sqrt{.2441} = .49$ seconds; for weight, $\sqrt{32.22} = 5.68$ pounds; and for IQ, $\sqrt{2.68} = 1.64$ points.

It is now possible to make some statement about the unknown population parameter and to make the statement with a probability of being correct by a specified amount. The estimated standard error has given information about the distance the means in the sampling distribution have fluctuated. Since the sampling distribution is a probability distribution, it is possible to determine the range within which these means will be found, say, 95 times in 100. An appropriate probability table[10] may be used to determine the number of standard errors above and below the mean that will include 95 percent of the means in the sampling distribution. The probability table of the t distribution in Walker and Lev (14:465) is such a table. It is the appropriate table to use when the sample is small. Under the column marked "$t_{.975}$" (the column to use in establishing the 95 percent confidence limits) and across from the $n = 9$, ($n - 1$ degrees of freedom), a value of 2.26 is found. Then, 2.26 standard errors above and below the obtained mean of the sample would establish *confidence limits* for the population mean. One would expect the mean of the population to be within these confidence limits 95 times out of 100. Five times out of 100, the true population mean would be outside these confidence limits. For the dash, for example, the true mean of the population would be 6.07 ± 2.26 (.49), or between 4.96 and 7.18 with a 95 percent probability of being correct. It may be noted that the mean of the population (see footnote 8, p. 82) is 5.90. Thus, the true mean (μ) is within the confidence limits. The reader may calculate confidence limits for weight and IQ and check to see whether the population mean is within the confidence limits.

A band, for the mean of the population, however, from 4.96 to 7.18 seconds seems rather wide for such an event as the 30-yard dash. Is there any way in which the standard error may be narrowed? The formula for estimation of the standard error ($s/\sqrt{n}$)

[10]Ordinarily the tables used are the normal probability tables such as may be found in Walker and Lev (14:456-57). Usually the normal probability table is used when the sample is of size 30 or larger. This is the table that indicates that 1.96 standard errors are to be used to establish 95 percent confidence limits.

in simple random sampling gives the clue. If the number (n) in the sample is increased, then the estimated standard error should decrease. To illustrate this point, another independent sample of size 40, n = 40, has been drawn from the same population. A summary of the results is given in Table 2.

For this sample of size 40, the sampling ratio is 40/103. The fpc is 1–fpc = 1–40/103, or .611. The reader may be interested to note that the sample means of the dash and weight for the sample of size 40 are closer to the population values than the sample of size 10. It is also interesting to note that the est. σ_M for all three measures has been reduced. The mean for the IQ of the sample of size 10 is closer to the mean of the population than the mean of the sample of size 40. But, the est. σ_M for IQ of the sample of size 40 is smaller than for the sample of size 10. The

TABLE 2.—SUMMARY OF SCORES FOR SIMPLE RANDOM SAMPLE, SIZE 40

Statistic	Formula	50-yd. Dash	Weight	IQ
Sum of raw scores	ΣX	230.10	[illegible],620.00	2,632.00
Sum of squared scores	ΣX^2	1,371.35	366,218.[illegible]	175,366.00
Sample mean	$\bar{X} = \frac{\Sigma X}{n}$	5.75	90.50	65.80
Sample variance	$s^2 = \frac{\Sigma X^2 - \bar{X}\Sigma X}{n-1}$	1.24	989.95	55.91
Sample standard deviation	$s = \sqrt{s^2}$	1.11	31.46	7.48
Estimated sampling variance	est. $\sigma_{\bar{x}}^2 = (1 - \text{fpc})s^2/n$	.19	15.14	1.40
Estimated standard error	est. $\sigma_{\bar{x}} = \sqrt{\text{est. } \sigma_{\bar{x}}^2}$	.43	3.89	1.18

95 percent confidence interval for the sample of size 10 ranges from 63.39 to 70.81; the sample of size 40 ranges from 63.49 to 68.11.[14] It may readily be seen that the range has been decreased by using the larger number. It is also interesting to note that the population mean (see footnote 8, p. 82) is within the confidence limits set.

One further comment should be made before leaving the simple random sampling discussion and moving on to other sampling

[14] It should be remembered that the t table in Walker and Lev (14:465) is appropriate for use with the sample of size 10. The normal probability table in Walker and Lev (14:456-57) was used for the sample of size 40.

designs. The effect of increasing the size of "n" has been demonstrated as one way to decrease the size of the standard error. There is yet another quantity in the formula s^2/n that should be considered in relation to the size of the sample selected. The value substituted in the formula for s^2 is taken from the variance of the sample. Therefore, if the variability of the characteristic is known to be large from age level to age level, or by whatever characteristic the population is classified (stratified), then the sample should be larger than if the variability is small. It is in these instances that different sampling designs are effective. It is possible to *design* (select and estimate) a sample that more adequately takes this difference in variability into account. The reader should be cautioned, however, that the same methods of estimation of the standard error *will not hold* when different designs are used. Other restrictions than the fpc are introduced in other designs and they must be properly accounted for in the formulas used.

Stratified Random Sampling. Stratifying the population will sometimes increase the precision of the estimated standard error. Before a sample is drawn, the population is divided into strata by some characteristic and a random sample is taken from each stratum. In order that the stratified design may be efficient, the units in each stratum should be as homogeneous as possible and have real differences between strata. The stratified sampling design will show much gain in precision when the characteristic under observation has some correlation with the characteristic by which the strata are defined. For instance, chronological age may be expected to have some relation to weight. Stratifying the population by age, then, should show gains in precision of the estimated standard error of the mean for weight.

In the example to follow, the population of 103 girls is to be stratified by chronological age. Before selecting the sample, it was decided that the total sample number should consist of 20 units (20 individuals chosen from the seven strata—8-, 9-, 10-, 11-, 12-, 13-, and 14-year-olds).[18] While the sample to be chosen is composed of 20 people, a real difference in the method of selection

[18]There is no rule that specifies that each age must be in a separate stratum. It is possible that some ages, such as the 8- and 9-year-olds, might be combined in one stratum. However, for simplicity, each age is to be kept in a separate stratum for this example.

is to be encountered with the stratified design. In the two illustrative simple random samples, the entire sample was chosen before stopping. Not so in the case of the stratified sample. The "at random" feature of selection remains the same, but rather than selecting all 20 individuals in one drawing, the 20 individuals are selected in seven different and independent selections, one random sample from each stratum.

The question of how many persons shall be used in a stratum immediately is apparent. How shall one divide the 20 people to be selected in the sample among the seven strata? Part of the answer to this question lies in the number of people in each stratum.[16] Sometimes there is no advance information about the number of units in the population in each stratum. In such a case, some estimation must be made. Sometimes an authority in the area can give some very useful estimates. Another method would be to take a simple random pilot survey and estimate the number in each stratum. In many experiments, however, this does not present a problem since the sample is usually selected from some population that is fairly well enumerated.

In the population of 103 girls, the number in each stratum is known. As may be seen in Table 3, there are 12, 22, 13, 12, 15, 15, and 14 girls respectively in strata one through seven. One way to determine the number in the sample for each stratum is to calculate the proportion of the total population in each stratum.[17] Take for example the proportion of the total population that is in stratum one. Here, $N_h/N = 12/103$, or 12 percent of the total population is in this stratum. The .12 is the *weight* of the stratum. If the total size of the sample (n) is multiplied by the weight of the stratum (nw_h), the number of units to be chosen for the sample in that particular stratum is determined. Since $20 \times .12 = 2.4$, two people will be randomly selected from the 12 for the sample in stratum one. The same procedure is followed for each stratum. The sum of the proportions for each stratum must equal 1.0 or $\Sigma w_h = 1$. This procedure gives a *proportionate* stratified sample. It has an advantage of simpler calculational procedure than other methods of choosing the number sampled in each stratum.

[16] Theory suggests that if a few strata are greatly skewed (strata that "tip the scales" in one direction) then all units in such strata should be used in the sample. Appropriate formulas and procedures will be found in sampling texts.

[17] Sometimes the proportions in strata are deliberately made unequal. Then, the weight of the stratum must be used in all calculations.

There is yet another way in which one may look at the size of the sample in each stratum when selecting a proportionate stratified sample. The sampling fraction, n/N, will give the proportion of units to select in each stratum. In this particular example, $20/103 = .19$, or 19 percent of the number in each stratum should be selected for the sample. If .19 is multiplied by the number in each stratum (N_h), then the size of sample for each stratum may be determined. The reader will note that either procedure will give the same number for the sample in each stratum.

Table 3 is presented for purposes of helping the reader become familiar with the calculational procedures used in stratified random sampling when a proportionate sample has been drawn. Again, the sample was drawn by selecting a random sample in each stratum.[18]

As may be seen in Table 3, means, estimated variances, and estimated standard errors have been calculated for each stratum. The problem now is one of combining the data for each stratum in some fashion to get the sample mean. With proportionate sampling, the calculation of the mean for the sample is a simple matter. The total sum of the sum of scores for each stratum (see Table 3, column marked "Total") may be divided by the number of people in the sample; thus $\bar{X}_{st}$[19] $= 1/n\ (\Sigma X)$. Means for the sample on the three items, then, are dash, $114.4/20 = 5.72$ seconds; weight, $1678/20 = 83.90$ pounds; and IQ $1293/20 = 64.65$ points.

If, on the other hand, the number in each stratum had not been proportionate, each stratum mean would have to be multiplied by the stratum weight before summing. As an illustration, the procedure for estimating the sample mean for the dash would be as follows:

$$\bar{X}_{st} = \Sigma w_h \bar{X}_h, \text{ or } \bar{X}_{st} = (.12 \times 6.55) + (.21 \times 6.75) + (.13 \times 5.70) + (.12 \times 5.60) + (.14 \times 5.33) + (.14 \times 5.03) + (.13 \times 4.97) = 5.713$$

This *weighted* procedure has given the same value as the *proportionate* procedure for the mean, except for rounding errors. The reader is cautioned that a weighted procedure is necessary in all

[18] The reader will note in Table 3, as well as in Table 5, that formulas are given only for the data on the 30-yard dash. They are not repeated for data on weight and IQ since the same formulas apply.

[19] $\bar{X}_{st}$ is the symbol used to represent the sample means for a stratified sample.

TABLE 3.—SUMMARY DATA FOR STRATIFIED SAMPLE

Statistic	Symbols	8 yrs.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs.	14 yrs.	Total
Stratum number		1	2	3	4	5	6	7	
Number in population	N_h	12	22	13	12	15	15	14	103
Stratum weight (N_h/N)	w_h	.1165	.2135	.1262	.1165	.1457	.1457	.1359	1.00
Number to be selected	$n_h w_h$	2	4	3	2	3	3	3	[illegible]
Dash									
Sum of scores	ΣX	13.1	27.0	17.1	11.2	16.0	15.1	14.9	114.40
Sum of squared scores	ΣX^2	86.21	185.08	97.97	63.04	85.84	76.13	74.53	668.80
Variance in sample	$s^2 = \frac{\Sigma X_i^2 - \bar{X}_i \Sigma X_i}{n_i - 1}$	.40	.91	.25	.32	[illegible]	.09	.24	
Mean	$\bar{X}_h = \frac{1}{n_h}\Sigma X_h$	6.55	6.75	5.70	5.60	5.33	5.03	4.97	
Est. variance of $\bar{X}_h$	est. $\sigma_{\bar{x}_h}^2 = \left(1 - \frac{n_h}{N}\right)\frac{s_h^2}{n_h}$	.15	.18	.07	.13	.08	.02	.06	
Est. standard error	est. $\sigma_{\bar{x}_h} = \sqrt{\text{est. } \sigma_{\bar{x}_h}^2}$	.40	.43	.26	.36	.27	.15	.25	
Weight									
Sum of scores		125.00	252.00	220.00	182.00	33[illegible].00	268.00	296.00	1,678.00
Sum of squared scores		7,897.00	16,374.00	16,366.00	17,012.00	39,731.00	24,458.00	29,546.00	151,384.00
Variance in sample		84.50	166.00	117.20	450.00	1,162.45	258.78	171.32	
Mean		62.50	63.00	73.33	91.00	111.66	89.33	98.66	
Est. variance of $\bar{X}_h$		34.05	33.44	31.48	181.31	312.23	69.51	46.02	
Est. standard error		5.83	5.78	5.61	13.46	17.67	8.49	6.78	
IQ									
Sum of scores		128.00	255.00	217.00	113.00	206.00	179.00	195.00	1,293.00
Sum of squared scores		8,194.00	16,451.00	15,801.00	6,445.00	14,230.00	10,683.00	12,699.00	84,493.00
Variance in sample		2.00	64.92	52.70	60.50	43.02	1.93	7.00	
Mean		64.00	63.75	72.33	56.50	68.66	59.66	65.00	
Est. variance of $\bar{X}_h$		.81	13.08	14.15	24.38	11.56	.52	1.88	
Est. standard error		.90	3.62	3.76	4.94	3.40	.72	1.37	

stratified samples that do not have proportionate subsamples, i.e., are self-weighting.

The estimated variance of the means of the subsamples may not be summed as simply as the scores were to determine the mean of the sample. Each stratum estimated variance (est. σ_h^2) must be multiplied by the square of the stratum weight (w_h^2) before summing. The formula for the sample variance of means in stratified sampling then, is

$$\text{est. } \sigma_{\bar{x}_{st}}^2 = \Sigma w_h^2 \, (\text{est. } \sigma_h^2).$$

Substituting the appropriate values for the dash from the table into the formula, we have est.σ_{st}^2 = $(.12^2 \times .16) + (.21^2 \times .18) + (.13^2 \times .07) + (.12^2 \times .13) + (.14^2 \times .08) + (.14^2 \times .02) + (.13^2 \times .06) = .016270$. The standard error is

$$\text{est.}\sigma_{st} = \sqrt{.016270}, \text{ or } .13.$$

The estimated variances and standard errors of the means for weight and IQ are

	est. σ_{st}^2	est.σ_{st}
Weight	13.3677	3.66
IQ	1.447	1.20

Again, as in the case of the simple random sample, with this information, confidence limits may be set for the mean of the population. Further examples will not be given here, but the reader may wish to set the confidence limits and compare them with the simple random sample confidence limits.

For the purpose of comparison, suppose we examine the estimated standard errors calculated from the three samples. They are compiled in Table 4. Note the size of the standard error for the simple random sample of size 40 and for the stratified sample of size 20. For the dash, the size of the standard error has noticeably been reduced by stratification, even though the stratified sample has just half as many individuals in the sample. Also for weight, the standard error is smaller than for the simple random sample of size 40. Running skill and weight are both correlated with age, the characteristic by which the strata were defined. In general, the stratification technique will show gains in precision if the characteristic under consideration is related to the characteristic by which the strata are defined. On the other hand, note the standard error for the mean of the IQ. For all practical purposes, the standard error for the simple random sample of size 40 and the stratified sample of size 20 are the same. Actually one does not expect the IQ to change as the child grows older. Consequently one would not expect a gain in precision in estimating the average IQ from stratification by age.

TABLE 4.—SUMMARY OF ESTIMATED STANDARD ERRORS

Type sample	n	Dash	Weight	IQ
Simple random	10	.49	5.68	1.64
Simple random	40	.43	3.89	1.18
Stratified random	20	.13	3.66	1.20

Cluster Sampling. Individuals are not selected at random in cluster sampling. Rather, the sampling unit is a "bunch" or cluster of elements. Quite frequently when the population inhabits a large area, collecting data from a simple random or stratified sample results in much time spent in traveling at a considerable cost. Under such circumstances, the use of a cluster sample may be considered. Some precision may be lost by cluster sampling, but time may be gained and cost reduced. The cluster sampling technique also lends itself well to experimentation in schools where whole classes are taken. The reader must be cautioned, however, that the practice of taking a "convenient" class does not come under the heading of "probability sampling." The clusters must be chosen at random.

Cluster sampling is different from stratified sampling in yet another way. In stratified sampling one attempts to define strata in such a way that all elements in each stratum are as nearly alike as possible. In cluster sampling, the more heterogeneity within a cluster the better.

The researcher does not determine the exact number of people to be selected in a cluster sample, since the number of people in the cluster probably cannot be controlled. Very seldom in practice will one be able to randomly select clusters that have an equal number of units. The calculational procedure would be greatly simplified if such were the case, however.

Let us return again to the population of 103 girls for an illustration of cluster sampling. It was decided that each school would constitute a cluster. All 13 schools were listed in alphabetical order and numbered in order from 1 to 13. It was decided that three clusters would be drawn. A table of random numbers was used to learn that clusters 4, 7, and 10 should be the sample. Scores for every individual in each cluster were summed. The data are summarized in Table 5.

TABLE 5.—SUMMARY OF SCORES FOR THE CLUSTER SAMPLE

Statistic	Symbols and formula	Clusters 1	2	3	Total
Number in each cluster	n_i	14.00	15.00	3.00	32.00
Average number in cluster	$w_i = n_i/\bar{n}$	1.31	1.41	.28	
For dash					
Sum of scores	ΣX_i	81.50	86.00	17.80	
Sum of squared scores	ΣX_i^2	484.61	501.72	105.70	
Mean for each cluster	$\bar{X}_i = \Sigma X_i/n_i$	5.82	5.73	5.93	
For weight					
Sum of scores		1,278.00	1,288.00	165.00	
Sum of squared scores		131,692.00	115,852.00	9,161.00	
Mean for each cluster		91.29	85.87	55.00	
For IQ					
Sum of scores		955.00	1,014.00	199.00	
Sum of squared scores		65,895.00	70,206.00	13,289.00	
Mean for each cluster		68.21	67.60	66.33	

It may be noted that the three clusters have 14, 15, and 3 individuals respectively. When the numbers in each cluster are not equal, a *weighted* method of combining the means of each cluster is used to determine the sample mean. In order to get the weight of the mean, however, the average number ($\bar{n}$) in each cluster must be calculated. Let "m" equal the number of clusters in the sample. ("M" equals the number of clusters in the population.) Then, if the number (n) of individuals in the sample is divided by the number of clusters in the sample ($n/m = \bar{n}$), the average number in each cluster may be ascertained. In this example, the average number in each cluster is 32/3 = 10.67. Now, if the number in each cluster (n_i) is divided by the average number ($\bar{n}$), the weight of the cluster is determined ($w_c = n_i/\bar{n}$). The weight of each cluster must be multiplied by the mean of each cluster before the cluster means may be summed and averaged.

This operation is symbolized[20] as $\bar{X}_c = \frac{1}{m}\sum^{m}\frac{n_i}{\bar{n}}\bar{X}_i$. Suppose the

[20] $\bar{X}_c$ is the symbol to be used for the sample mean of cluster sampling. If all cluster means are summed by a weighted ($n_i/\bar{n}$) procedure, and divided by the number of clusters (1/m), this average is the cluster sample mean.

appropriate values for the dash are substituted in this formula. Then we have $\bar{X}_c = 1/3\ [(5.82 \times 1.31) + (5.73 \times 1.41) + (5.93 \times .28)] = 5.78$. (Other cluster sample means are weight $= 85.36$; IQ $= 67.75$.)

A comment should be made prior to demonstrating the estimation of the variance and the standard error of the mean. As will be recalled, a *finite population correction* (fpc) term was used in the other samples. It is to be used here also. Since clusters were drawn rather than individuals, however, the fpc term becomes m/M rather than the n/N as used with the other samples. The formula for calculating the estimated variance is

$$\text{est. } \sigma_{M_c}^2 = (1 - m/M)\ 1/m \left[1/(m-1) \sum^{m} \left(\frac{n_i}{\bar{n}}\right)^2 \left(\bar{X}_i - \bar{X}_c\right)^2 \right]$$

The student should note that the sample mean ($\bar{X}_c$) is subtracted from each cluster mean ($\bar{X}_i$) in the formula. This is a different procedure from that encountered in stratified sampling. Such a procedure introduces the notion of variance of the means in the sample.

Now, let us substitute the appropriate values for the dash into the formula.

$$\text{est. } \sigma_{M_c}^2 = (1 - 3/13)\ 1/3 \left[\frac{(1.31^2 \times .04^2) + (1.41^2 \times .05^2) + (.28 \times .15^2)}{3 - 1}\right] = .00121344$$

The estimated standard error is est. $\sigma_{M_c} = \sqrt{\text{est. } \sigma_{M_c}^2}$, or .0348. The estimated variances and standard errors for the other two measures are

	est. $\sigma_{M_c}^2$	est. σ_{M_c}
Weight	17.0168	4.13
IQ	.0720	.27

Suppose we compare the estimated standard errors calculated from the cluster sample with the estimated standard errors from the simple random and the stratified samples. Referring to Table 4, the reader will note that the cluster sample estimated standard errors for the dash and the IQ are smaller than in the other sampling designs. One does not ordinarily expect cluster sampling to give a more precise estimate than the simple random or the stratified design. One expects to lose some precision when using the cluster design. Such is the case with the estimated standard error for weight which was 4.13. As can be seen, the estimated standard error for weight with a sample of 30 individuals is larger than the stratified sample with 20 individuals.

Again, confidence limits for the true mean may be calculated. The reader will discover that the mean of the population for the dash is outside the 95 percent confidence limits. This can happen, since a 95 percent confidence interval indicates that 5 times in 100 the true mean would be outside the confidence limits.

Other Sampling Designs. The three most commonly used sampling techniques are the designs that were discussed in the preceding section. There are other sampling designs, however, that are sometimes encountered in the literature. Each of the designs has its own formulas, as was the case with the three designs illustrated above. Again, the reader is cautioned that very misleading results may be reported if the researcher is not careful to use the formulas that are specific for the design. Brief descriptions of some of the other designs will be given below. Examples will not be given, nor the correct formulas. The reader may find the correct procedure to follow in selection, and the appropriate formulas to use in estimation in the sampling texts referred to previously.

Systematic Sampling. This is a technique that is used quite often when a large card file or some large list is available. The procedure is to decide upon how many units are wanted in the population. The number in the sample is then divided into the number of units in the population to ascertain the number to be used when selecting the "beginning" point in the file or list. Suppose for instance that there are some 15,000 children in a city school system and that some list or file is available. Further, suppose that some 500 children are to be included in the sample. Then, $15{,}000/500 = 30$, the number with which the researcher is concerned when entering the table of random numbers. He enters the table of random numbers to locate the point at which he is to begin sampling between 1 and 30. Suppose further that the first number encountered between 1 and 30 happens to be 7. No other number is necessary, since every 30th card thereafter is drawn. Then, the researcher goes to the file and selects card 7, 37, 67 . . . and so on, until all cards have been canvassed.

The researcher who employs this sampling design should be alert to two departures from randomness that are sometimes found in a file of cards. One such departure is called a *trend*. Suppose, for example, that the researcher is interested in the average age of employees in a large factory. And suppose that someone in the office has arranged the cards in the file by age in years and months. Then the estimated mean of the population would differ markedly from sample to sample depending upon the number selected at random for the first draw. It is easy to see that there might be great differences in the ages recorded from a sample consisting of cards 2, 32, 62 . . . 902 and a sample consisting of cards 29, 59, 89 . . . 929. Another departure from randomness sometimes found in card files is spoken of as *cyclical fluctuation* or *periodic variation*. Such a fluctuation would be found if, say, there were 15 employees in each department of the factory and the card file had been arranged by department from youngest to oldest. Again, one would find wide variability in samples depending upon the starting point.

Multi-Stage Sampling. This is a type of sampling procedure that is done in more than one stage. Usually a large, not too comprehensive, survey is made. After examining the results of the first survey, another sample is selected. The second stage of sampling is apt to be a more comprehensive investigation than the first. Samples may be selected in any number of stages and any type sampling design may be used at any stage. If the design does not go beyond the second stage, however, the design of the sample in the report is apt to be spoken of as a *double stage* sampling technique. The student who plans to use a multi-stage design should consult Cochran (1:215-67), Deming (6:135-65), and Hansen (7:366-424).

Sequential Analysis. This is a sampling design that is often applied in production where some decision must be made about selecting a box of bolts or rejecting the bolts—or any other manufactured objects. Much theory of sampling has been developed around this type of design. It is sometimes referred to as *acceptance sampling*. For a more thorough discussion of this sampling design, the reader should consult Deming (6:277-82).

Interpenetrating Replicate Subsamples. This sampling design was used originally to check one enumerator against another. It seems to be gaining in favor, however, and probably should be described here. Actually it is useful for other purposes. One of the main purposes for this design is to measure the degree of agreement between subsamples. In general, the procedure is one of selecting, say, ten subsamples from the total population. Units in the first set of ten subsamples are then further assigned at random to ten other subsamples in such a way that each of the new set of ten subsamples contains units from each of the first set. The reader should refer to Cochran (1:312-15), before using this sampling design. Jones (8) also has a useful explanation of this sampling design.

NONPROBABILITY SAMPLES

In the preceding sections, emphasis has been given to techniques of selecting an effective sample by random methods. Repeatedly the point has been made that randomness must be used at some stage in the selection of the sample if one is to be able to justify inferences about the population from which the sample is taken. This point cannot be too strongly emphasized, since such a procedure is absolutely essential if generalizations about the population are to be defended on the basis of probability theory. Any time a standard error term is used to infer real differences between groups or to state confidence limits for population parameters, a probability sample must be used.

It is unfortunate that researchers, because of lack of subjects or because of the "convenience" of certain groups, have utilized non-probability samples. Non-probability samples are those which were not selected at random, but rather by some other method of selection. It is not uncommon to find examples of such samples in the literature. One frequently may read a research report that has used "intact classes" for the sample. These particular "intact classes" might have been selected because they met at an hour when the researcher was free, or the instructors of the classes might have been willing to co-operate with the researcher. One sometimes reads of experiments where two groups are compared, or two or more methods are compared, and the intact groups have been used with no attention to assignment "at random" of subjects, of methods, or of instructors. Sometimes a researcher will use "volunteers" for the sample. Reports of research are also found in which a *quota* of subjects is selected by taking the number decided upon from individuals met in a hallway or on a street corner. Another practice has been one of selecting "typical" classes or persons. Such a practice places unwarranted responsibility on the individual who supposedly is capable of selecting "typical groups."

Much too frequently, reports using nonprobability samples summarize the data and make generalizations about the population, using probability tables in so doing, that cannot be defended. The methods of selecting the samples have not been chance methods. When a non-probability sample has been selected, Cochran (2) points out that interpretation of results depends more and more on the expert in a particular field, not on mathematical probability or on help from a mathematical statistician.

BIBLIOGRAPHY

1. Cochran, William G. *Sampling Techniques.* New York: John Wiley & Sons, 1953.
2. Cochran, William G.; Mosteller, Frederick; and Tukey, John W. "Principles of Sampling." *Journal of American Statistical Association* 49:13-35; 1954.
3. Cornell, Francis G. "Sample Surveys in Education." *Review of Educational Research* 24:359-74; 1954.
4. Cornfield, Jerome. "Modern Methods in the Sampling of Human Populations; the Determination of Sample Size." *American Journal of Public Health* 41: 654-61; 1951.
5. Deming, W. Edward. "On Simplifications of Sampling Design Through Replication with Equal Probabilities and Without Stages." *Journal of American Statistical Association* 51:24-53; 1956.
6. Deming, William E. *Some Theory of Sampling.* New York: John Wiley & Sons, 1950.
7. Hansen, Morris H.; Hurwitz, William N.; and Madow, William G. *Sample Survey Methods and Theory.* Vol. 1 and Vol. 2. New York: John Wiley & Sons, 1953.
8. Jones, Howard L. "The Application of Sampling Procedures to Business Operations." *Journal of American Statistical Association* 50:763-74; 1955.
9. Kish, Leslie. "Selection of the Sample." *Research Methods in the Behavioral Sciences.* (Edited by Leon Festinger and Daniel Ketz.) New York: Dryden Press, 1953.
10. Kish, L. "Confidence Intervals for Clustered Samples." *American Sociological Review* 22:154-65; 1957.
11. Moonan, William J. "On the Problem of Sample Size for Multi-Variate Simple Random Sampling." *Journal of Experimental Education* 22:285-88; 1954.
12. Slonim, Morris James. "Sampling in a Nutshell." *Journal of American Statistical Association* 52:145-61; 1957.
13. Stephan, F. F. "Advances in Survey Methods and Measurement Techniques." *Public Opinion Quarterly* 21:79-90; 1957.
14. Walker, Helen M., and Lev, Joseph. *Statistical Inference.* New York: Henry Holt and Co., 1953.
15. Yates, Frank. *Sampling Methods for Censuses and Surveys.* New York: Hafner Publishing Co., 1949.

CHAPTER 5

Tools for Obtaining Data

ESTHER FRENCH
CHARLES C. COWELL
ALFRED W. HUBBARD

AS A STUDY IS BEING PLANNED, THE RESEARCH WORKER IS FACED with the decision of *how* the solution is to be reached, *how* data will be obtained, *which* tools or devices are available and best suited to solve the problem. The research worker cannot proceed without "tools" of the job any more than a plumber can work without his tools, or a typist without her typewriter, or an artist without materials and equipment.

Furthermore, it is essential that exactly the right tool be selected for optimum results. Judgment for the selection is partly the result of knowledge about the tool—its function and limitations—and partly the result of experience in using it. The golfer with training and experience knows that a putter, an iron, and a wood are designed for different purposes. The fisherman knows which equipment and lure to use for different kinds of fish. The camera fan knows the difference in potential of the various lenses, shutter speeds, and films. Also, in each case, the performer or worker knows what is available for purchase, in anticipation of optimum results in the next endeavor.

There are many tools that have been developed for research purposes. Some have originated with education; others have been adapted to educational needs, having first appeared as tools for medical, scientific, or sociological research. The discussions of this chapter attempt to present briefly some of the tools which seem to be most useful to the research worker in the fields of health, physical education, and recreation. Attitude scales, socio·metric techniques, and photography are discussed in more detail than the typical research tools in education because of their special application to these fields. The specifications for each tool are pre·cise, usually based on earlier research. The lists of references will help the reader trace development and should be used to supple·ment the brief presentations in this chapter.

The investigator should always be on the alert for new modifica·tions of tools and for totally new devices. Also, this phase of re·search offers opportunity for a creative approach to problem solving.

Typical Research Tools in Education

ESTHER FRENCH

INTERVIEWS AND QUESTIONNAIRES

Interviews and questionnaires have much in common. Both are survey tools used for the purpose of obtaining data concerning present status, practices, or opinions regarding a selected situation or problem. The interview has been called an oral questionnaire; it has also been defined as a conversation with a purpose. Occa·sionally, the two have been employed in the same study, supple·menting each other. When this is done, the interview is used to secure the less factual data on information concerning matters that persons may consider too confidential to put into written form. Matters of personal habits, family life, attitudes, and beliefs lend themselves to the interview approach.

The questionnaire is more commonly used for quickly obtaining information from a large number of persons concerning factual matters. However, the interview can also be used for obtaining information from large numbers of persons, as has been demon·strated by Gallup and others engaged in the opinion poll business and also by Kinsey and his co-workers in their well-known studies of sexual behavior.

Whenever a question is asked—either orally, as in the interview, or in written form, as in the questionnaire—it may be subject to various interpretations. If the question is highly ambiguous, the replies may be colored. In an interview, it is usually possible to detect when a question is being misinterpreted and clarifications can be made. Another advantage of the interview is that frequently additional information relevant to the general problem is revealed.

It has been estimated that at least one-fourth of all the published educational studies have used the questionnaire technique. Misuse and overuse of the questionnaire have caused it to be criticized. As with any tool, the questionnaire must be properly used and the data obtained must be correctly interpreted, or the results are worthless. Its use should be limited to the types of studies for which data are available by no other means, and the results should be interpreted strictly within the limits of the facts.

The interview is difficult to use effectively unless the interviewer is skillful and well trained. The quality of the information obtained depends upon the quality of the interviewing. The interviewer must know his topic thoroughly and be adept at winning the confidence of the persons being interviewed. Consideration should be given not only to the phrasing of questions but to the timing of them.

The selection of subjects (respondents) to be surveyed greatly affects the results. This holds true regardless of the technique used. If expert opinion is wanted, then a few carefully selected authorities will provide better data than a large number of less qualified persons. The indiscriminate use of "big name" persons, however, is no guarantee of expert opinion as few persons are expert concerning all matters. If the problem is truly important and the data cannot be otherwise obtained, the majority of professionally minded persons will reply personally to a well-prepared questionnaire or will grant the courtesy of an interview. Recommended procedures for sampling should be studied and followed. The importance of using proper sampling cannot be overstressed and applies to all research studies, regardless of the selection of tools for obtaining data. Every care should be taken to ensure that the sample is unbiased. Size, alone, is no guarantee of freedom from bias.

Respondents should be asked only for information they can and will give. The director of recreation or of physical education in a large city system should not be expected to be intimately acquainted with small details of the operation of the swimming pool, playground, or gymnasium. The playground director or teacher, working in one situation, should not be expected to supply information on the over-all policies. Anonymity must be assured if personal questions are asked—as for example, questions regarding salary, age, or evaluations of the efficiency of co-workers or superiors.

Each question or item under consideration for inclusion should be evaluated as to its form and function. Criteria should be applied, such as the following: Exactly what is the item intended to measure? Does the question contribute to the solution of the problem? Is there ambiguity? Can it be made more clear? Are there any unnecessary qualifying phrases that might start the respondent to thinking along irrelevant lines? Is the question straightforward and direct?

Submitting the questions to a trial run both in written and oral form should aid in improving them. Rereading after a lapse of a few days is another good procedure. Care in preparation pays off not only in a higher percentage of returns but, what matters even more, in increased value of the obtained data. Thorough prestudy of the field permits better delimitation of the problem. Any questions which duplicate or overlap others excessively should be eliminated.

The form of the questionnaire should be such that responses can be made easily—by checking, by yes or no answers, or by very few words. The responses should lend themselves to tabulation. Items requiring lengthy responses cannot be readily tabulated. If it can be anticipated that qualified replies rather than "yes-no" replies are likely to be encountered, then the questionnaire should provide for these. This can be done in the same manner as for a rating form, with directions included. For example, in seeking an opinion, consideration might be given to including three columns, headed "Agree," "Disagree," "Uncertain."

A stimulating letter of transmittal; the prestige of a reputable sponsoring agent; the follow-up of returns; the inclusion of self-addressed, stamped envelopes for replies; and the offer to supply a summary of the results, if desired, are all a part of the procedures which have generally been found valuable. But the most important procedures for securing validity of data are concerned with care in the preparation of questions and the use of proper sampling techniques. Definition of terms is frequently necessary and generally the definitions are included immediately preceding the questions where the terms appear. Since the questionnaire has had such wide usage, there are several sources from which detailed and helpful information may be obtained (6, 13). The interview has had more limited use (3).

OBSERVATIONS, CHECKLISTS, RATING SCALES

The observational method as used by the research worker is a planned procedure directed toward seeing and noting the amount or degree of specific practices with relation to a definite problem. It is not mere visitation or "looking." Checklists, rating scales, and similar devices have been developed to facilitate the recording of data.

Observations. These have been used extensively in studying the behavior of children, the performance of student teachers, the rating of sports officials, and as a preliminary step in many research studies. Usually, it is done in a natural setting with no attempt made to elicit a specific response. One of the earliest studies making use of this technique is the now classic "Biography of a Baby," by Shinn, which reports a complete record of one child during the first year of life. The early form of recording was a diary or running account of all that took place, while the modern form consists of a checking scheme or rating scale. A checklist differs from a rating in that the number of occurrences or the presence or absence of a trait is recorded, without reference to a scale of values.

The use of observations is practically unlimited. Observations may be made of the process of learning as well as of the end results. The method has certain advantages. Frequently it requires no change of conditions and no apparatus. It does not require direct co-operation on the part of the persons being observed. It is particularly useful for determining reactions under customary conditions.

A high degree of objectivity in making observations is necessary if the data are to be valid, and objectivity is not easy to obtain. For example, three persons, viewing a ball rolling near a line, may differ in their judgment about its location when the whistle sounded. One may think that the ball was in the field of play, another that it was on the line, and the third may be equally positive that it was outside the field of play. The observers are more likely to agree if they view the situation from the same angle and distance, have equally good eyesight, and are devoid of any emotions concerning the outcome of the occurrence. Observers must be thoroughly familiar with the general field of study, and the specific problems being studied must be carefully defined. Both skill and accuracy in detecting pertinent factors are needed. If only one observer is to be used in collecting data, care should be taken to see that he is highly qualified.

Several observers should make concurrent observations, working independently, and the degree of agreement among judges should be checked. Unless the correlation is high, they should review the actions being observed, perhaps redefine them, and repeat the observations and procedures until high agreement is reached. Various types of training devices such as audio-visual aids, have been used to ensure higher agreement among judges. When it is not practical to give the observers extensive training, it may be preferable to rely on several judges rather than one and to use the sum of their scores. When the act to he judged needs to be viewed from several different aspects, as is the case in competitive diving, then several observers are used, even though high agreement among the judges has been obtained.

It is frequently desirable to give greater weight to certain phases of the problem than to others. For example, in rating badminton players, use of a variety of strokes may have more or less importance than ability to cover the court. In evaluating sanitary practices, the wearing of hair nets by food handlers might have a different value or weighting than the method used in sterilizing dishes.

As in all research, conclusions must not be drawn upon too limited or a biased sampling. The length, frequency, and variety of observations should be sufficient to reveal the true facts. The number needed varies from one study to another. There has been some evidence presented favoring an accumulation of a large number of very brief observations over a few continuous observations, but again, logic should be used in setting up the study, with the purpose kept clearly in mind. The number or amount of observations needed is usually determined by the point at which no new items are being observed or beyond which point the ratio between items remains more or less stable.

Records of activity may be observed by mechanical means, such as movie cameras, dictaphones, electrical eye counters. They may also be obtained stenographically. Symbols or codes are frequently used on charts when frequency and spatial relationships are being observed. The charts used by coaches and scouts are illustrative. Usually these are prepared and used as an aid in analyzing play or successful performances.

Accessibility of the observers may be a limiting factor. Various environmental factors also affect results, and every care should be taken to see that these are "normal" or controlled. Again, they vary with the problem or situation being observed. Some general examples are the time of day, the humidity and temperature, the events immediately preceding the observation, and outside or unusual distractions. The observations should be arranged at a time when it is anticipated that the thing or things to be observed will occur. For example, in a study of the frequency of use of various pieces of play apparatus, the observation should be made when the play apparatus is available for use and the children are free to use it. The number using each piece can be recorded, with various observers focusing their attention on specified pieces of apparatus. What constitutes "use" will have to be defined in advance. If the observation takes place on an unusually hot day or an unusually cold day, it is conceivable that the results may be affected. An injury or an unusually good performance during the previous play period might affect use of a certain piece of apparatus. Repeated short observations are thought to be less subject to chance fluctuation than is a single longer observation.

At times, certain modifications of external conditions are made to economize on time or to facilitate observations. For example, one-way-vision screens have been used to eliminate the consciousness that one is being observed. Closed circuit television is a modification of this device but is much more expensive.

Great care has been taken in some studies to ensure the reliability of the observer, but in many studies little, if any, attempt has been made to control the various factors known to affect behavior, attitudes, and reactions. In using the observation method for case studies, better results are likely to be obtained if some controls are established. For example, a child may react quite differently when in one situation than in another. If it is thought that his companions affect his behavior or reactions, then he should either be observed in several different situations with the same companions present or be observed in the same situation several times with different groups of companions. Only a fine line of distinction can be made between a controlled observation and the experimental method. If it is desired that check studies or longitudinal studies be made, some controls must be established.

Checklists. Forms used to record the number of times a certain event occurs or the presence or absence of a clearly defined trait or situation are called checklists. Their value is highly dependent upon the observer's ability to be objective and the quality of his judgment.

One example of the use of a checklist is in comparing the effect of rules changes on the number of times a particular skill or specified bit of strategy is used. For example, one might study the number of times a spike is attempted in volleyball in the nonrotation game as compared to the rotation game. Data shown are selected from that obtained on one official volleyball game and are presented to illustrate this use of a checklist or incidence chart.

Skill	*Incidence During Each Half*		*Total Times Used*
	Rotation	Nonrotation	
Serve	21	30	51
Pass	26	16	42
Spike	16	12	28
Block	0	0	0

Ratings. These are essentially directed observations. They have been used for a number of purposes. Perhaps one of the best known uses is as a substitute for objective tests when the latter are not available. Ratings are also used to supplement these tests. In the area of dance activities, for example, subjective ratings of performance are in common use. When form in performance is being judged (as in gymnastic meets, for example), complete reliance is placed upon a rating scale, involving subjective judgment, but the observer is directed and uses an agreed-upon procedure and weighting scheme.

Raters are limited in the accuracy of their ratings by their experience, by the opportunity provided for observations, by knowledge of the activity or trait being rated, by the degree to which they can be objective in making judgments, and by the ability to concentrate on the task at hand. The reliability of ratings can be increased by combining the ratings made by several judges of the same pupil or subject.

There are many types of rating forms. The graphic device or scale is illustrated below.

Superior	Above Average	Average	Below Average	Poor

Usually this is placed beneath a statement describing the trait or characteristic. In the place of descriptive terms, a numerical scheme may be used.

Sometimes it may be preferable to use a three-point scale such as "Agree, Disagree, Uncertain." This has had frequent use in attitude scales.

Regardless of the form used, the directions should be clear. For example, is each individual to be rated in relation to the group or in relation to an ideal?

In determining the number of categories, some consideration should be given to the degrees of distinction possible. If the opportunity for making observations or judgments is quite limited, fewer categories should be listed. The use to be made of the ratings is another factor. For example, if a group of performers in an activity are being rated and it is desired that ratings result in five or more ability groupings, a seven-point scale should be considered. This is assuming that several judges are being used and that they tend to avoid the extremes in recording their judgments. If just three groupings are needed, a five-point scale should suffice to give the desired spread.

Certain preliminary preparations can be made to increase the validity of the ratings. These include (a) determining the nature of the content of the activity or trait to be rated; (b) determining the number of categories to be used; (c) defining each category or point on the scale; (d) preparing the rating forms or score sheets in advance; and (c) selecting the raters and training them. The actual conduct of the ratings should be well planned to provide judges with the best views, sufficient amount of time, and freedom from distractions.

Ratings are of no value unless the observations upon which they are based are accurate. Unfortunately, persons confronted with a neat and inviting scale may be tempted to "co-operate" by putting checks in the spaces provided even though they have only a meager basis for judgment. In recognition of this, some forms provide a space for checking if you have had an inadequate basis for judgment. Amount of data and quality of data are two different things. The mechanics of filling out a rating form are easy; accurate judgments are difficult.

DOCUMENTS AND RECORDS

Documentary analysis is the study of a collection of written or printed materials to determine the frequency and usage of selected items or to reveal facts concerning an enterprise.

The kinds and sources of documentary data are varied. The more common types are administrative records, forms and reports; curricular materials such as syllabuses, courses of study, texts, notebooks, term papers, class papers and reports; legal acts and case reports; correspondence; and official reports on a governmental or institutional operation. There is an implication of veracity in the nature of the records. It does not follow, however, that the facts are pertinent to the study or in such form that they can be compared with other data.

The purpose of the study will determine the sources and kinds of information collected and selected, but the value of the information will depend upon how pertinent the items are in throwing light upon that purpose. The investigator should organize criteria into a rating scale or checklist appropriate to his study and sources. In general, the literature reveals the following criteria for items of documentary analysis: They should

1. Be valid and authentic
2. Be reliable and accurate
3. Be objective and carefully defined
4. Be representative, wide, and comparable samples
5. State limits, meaning, organization, and mechanical features
6. Be important and appropriate in terms of dates
7. Be feasible in terms of time.

Preceding the study of documentary data, the investigator should have an interview with the person in charge of the records or materials. Teachers, administrators, and supervisors especially need to be acquainted with textbook analyses. Objectives, outcomes, word counts, line or space counts, frequency of mention, importance, extent of use, and analysis of common errors exemplify helpful types of data and information in most fields; grade placements, difficulty of games, rhythms, stunts, or athletic events are often implied from course of study and textbook analysis.

Legal and official documents and records are preferred general sources. The study of daily notes of students or stenographic notes may be valid curricular sources. Syllabus and textbook analyses are reasonably reliable, but catalogue analyses provide the least efficient check. Official reports may give a better picture than exists, because of the desire to appear to meet standards and requirements.

The first step in a documentary analysis is to decide the kind of items needed. The formation of appropriate categories for collecting the data is a crucial problem. The data and information should be grouped so as to bring out differences, similarities, and functional relationships. Usually a small sample is used first and a trial classification is made, based upon reading and experience. A functional rather than a logical arrangement is the aim. The purpose of the study is the guide. Knowing the form in which data ultimately will be arranged enables one to get his basic data and information in order. Microfilming or other inexpensive reproduction of rare or otherwise inaccessible data is available in most libraries or institutions.

Interpretation follows the same rules as for other forms of research. Care needs to be taken that conclusions drawn from tabulations are justified by the relationships and representativeness of the data. The investigator's interpretations should be kept separate from the actual data. Findings should be checked against other reliable sources.

SCORE CARDS

Score cards have been used in appraising facilities, instructional and recreational programs, educational qualifications of teachers, and in connection with accreditation. The steps involved include the determination of scope; the establishment of objectives or criteria; the formulation of statements describing the various items; preparation of directions and forms; and selection of persons, schools, or localities to be scored.

One example is the "Criteria for Appraisal of the Instructional Programs of Physical Education in Colleges and Universities" which resulted from a Washington Conference (1). It was designed to serve as a convenient tool for program appraisal by administrative and faculty personnel in departments of physical education. It consists of 50 items, grouped under the headings of Philosophy and Objectives, Administration, and Program. Categories are provided on a five-point scale: Completely—5; To a great degree—4; To a moderate degree—3; Very little—2; Not at all—1. Two items are given here:

> 29. The activities selected make full use of local geography and climate.
> 30. The program provided opportunities through coeducational classes for teaching men and women to develop skills and enjoy together those activities which bring lifelong leisure-time satisfaction (1:37).

This is an example of a self-rating type. The LaPorte Score Cards (11) are an example of ratings to be made by outsiders. The Illinois Curriculum Program Study makes still another use of score cards. Their procedures involve ratings by parents and pupils as well as by teachers.

CRITICAL INCIDENT TECHNIQUE

The critical incident technique is a set of procedures used for collecting and classifying specific and significant behavioral acts taking place in defined situations. The primary purpose of the technique is to set up critical requirements for the performance of a specific activity or job. Flanagan (7, 8) and co-workers used it in 1940, in connection with their work in the Aviation Psychology Program, to evaluate behavior and performance of Air Force personnel.

Flanagan indicates the various uses: (a) to measure and evaluate actual performance; (b) to measure proficiency in sample situations; (c) to guide and organize training programs; (d) as a basis for selection and classification of workers, and (e) as a method of analyzing attitudes.

As a research tool for obtaining data, the critical incident technique has not been widely used in education but is included here because it appears to have value. Jensen (10) tried to answer the question, "What teacher traits produce effective teaching?," by collecting and recording first-hand reports of especially effective and ineffective teacher performance. He used the interview to obtain these reports. The reports of effective and ineffective performance were classified under major headings of personal, professional, and social qualities, and in turn under subheadings. From this classification, a list of critical requirements for teachers was constructed which formed the basis for evaluation of actual teacher performance in the Los Angeles area.

Using the critical incident technique as a research tool involves five procedural steps:

1. Determination of the general aim of the activity to be evaluated so that incidents may be observed in light of this aim
2. Development of specific instructions to the observers as to the situation to be observed
3. Collection of the data. This can be done through observation or it can be a collection of recorded data, secured by interview or questionnaire.
4. Analysis of the data for the purpose of summarizing and describing the collected observations
5. Interpretation of results within the limitations of the data and the technique.

There are obvious limitations to this technique, such as the difficulty in measuring the reliability and validity of the data collected by this means. However, the data might give a clue to the critical requirements, and tests might then be developed to measure these proficiencies or aptitudes.

SKILL TESTS

Criteria to be used in evaluating skill tests have been presented in several of the textbooks concerned with tests and measurements in the field of physical education. These include various statistical measures supplemented by logic and such practical considerations as economy of time and the availability of norms. When concerned with the selection of a skill test which is to be used as a tool for obtaining data for research purposes, the choice should be limited to tests with proven validity and reliability.

Validity is usually reported by means of a correlation coefficient expressing the degree of relationship between the test score and a criterion. If the criterion is a good one, then the higher the relationship, the more truly does the test appear to be measuring the ability in question. For many skill tests in sports activities, the criterion used has been a rating of general playing ability for the sport in question. Frequently, this results in a lower validity coefficient than would be obtained if the criterion were another test, proven to be valid, of the specific skill. For example, one should expect a lower relationship between a rating of general playing ability in tennis and a skill test of a single stroke, such as the serve, than one should expect between two skill tests, both assumed to be valid and reliable. If the skill test has high reliability and low validity, the nature of the criterion should be taken into consideration.

Validity should not be determined by statistics alone, regardless of the size of the correlation coefficient. Other considerations to be made in selecting a skill test for the purpose of obtaining data for research purposes include the following:

1. The test should measure an important ability.
2. It should involve one performer only.
3. It should provide accurate scoring.
4. It should provide a sufficient number of trials.
5. It should be of suitable difficulty.

Reliability is a prerequisite of validity, i.e., if the test is not consistent in its measurement of a given ability, it cannot be consistent in measurement of that ability represented by the criterion. The primary factors in obtaining consistent scores are adequate trials and objective scoring.

Reliability is expressed as a correlation coefficient. The calculation is ideally between scores obtained on two administrations on successive days. Practically, it is not always possible to administer the test twice, so the coefficient may be calculated on sums of alternate halves when the test has several trials.

KNOWLEDGE TESTS

In selecting a knowledge test to be used as a tool for obtaining data for research purposes, validity is the prime consideration. Does the test measure what it purports to measure or is it a measure of intelligence, of guessing ability, of memorization? Validity may be considered from a subjective, as well as an objective, viewpoint. Subjectively, the worth of a test may be evaluated by applying criteria such as the following:

1. Is the emphasis on functional value rather than on memorization of subject matter?
2. Is the test sufficiently comprehensive? Are any of the important outcomes of instruction ignored in the test as a whole?
3. Are the questions clearly stated?
4. Will it provide a wide range of scores, with no undue massing of scores at any one point?
5. Is the test sufficiently long to eliminate chance factors and yet not so lengthy that the slow readers are penalized?

Another method of checking the curricular validity of a test is to compare it against a carefully prepared outline of the unit or course.

An estimate of the validity of a knowledge test can be obtained by correlating the scores made on it with the test scores made by the same individuals on a previously validated test covering the same materials, should such a test be available. More frequently, a test is "validated" item by item against the criterion of the total score (number of correct responses) made by each person on that same test. If the test provides a wide range of scores and the items are skillfully prepared, then the use of the total score as a criterion is defensible. The analysis should reveal such information as the difficulty rating of the question (expressed in terms of the percent who succeeded in answering the question correctly), the functioning of the various parts (percent selecting each failing or incorrect answer), and the discriminatory power of the question. A question is said to have perfect discriminatory power when every student who answers the question correctly ranks higher on the total score scale than all students who answer it incorrectly. A question on which more students of low ability succeed than do students of high ability is said to have negative discriminatory power and is a poor question. Various methods have been devised for determining the discriminatory power of questions. See Chapter 8.

If a knowledge test is truly good, it should measure the ability of the student to make applications. The student possessing a considerable amount of knowledge should be able to apply facts learned to the solution of a new problem. However, if all the questions require application of knowledge, there is likely to be an undue massing of low scores since, in many classes, some of the students will succeed only on those questions covering content that has been memorized. The test should provide for as wide a range of abilities as are present in the group.

REFERENCES ON TYPICAL RESEARCH TOOLS IN EDUCATION

1. AMERICAN ASSOCIATION FOR HEALTH, PHYSICAL EDUCATION, AND RECREATION. *Physical Education for College Men and Women.* Washington Conference Report. Washington, D. C.: the Association, a department of the National Education Association. 1954. Revised, with additional recommendations, 1959.
2. BARR, ARVIL; DAVIS, ROBERT; and JOHNSON, PALMER. *Educational Research and Appraisal.* New York: J. B. Lippincott Co., 1953.
3. BINGHAM, W. V., and MOORE, B. V. *How To Interview.* Fourth revised edition, prepared with collaboration of John W. Gustad. New York: Harper & Brothers, 1959.
4. BLANK, LANE B. "Critical Incidents in the Behavior of Secondary School Physical Education Instructors." *Research Quarterly* 29:1-6; March 1958.
5. CUNNINGHAM and ASSOCIATES. *Understanding Group Behavior of Boys and Girls.* New York: Bureau of Publications, Teachers College, Columbia University, 1951.
6. DAVIS, ELWOOD C. "A Criticism of Survey Techniques in Health and Physical Education." *Research Quarterly* 12:27-36; May 1931.
7. FLANAGAN, JOHN C. "The Critical Incident Technique." *Psychological Bulletin,* 51:327-58; 1954.
8. FLANAGAN, JOHN C. *Flanagan Aptitude Classification Tests.* Chicago: Science Research Associates, 1953.
9. JAHODA, MARIE; DEUTSCHE, MORTON; and COOK, STUART. *Research Methods in Social Relations.* Part I. New York: The Dryden Press, 1951.
10. JENSEN, ALFRED C. "Determining the Critical Requirements for Teachers." *Journal of Experimental Education* 20:79-86; 1951.
11. LAPORTE, WILLIAM RALPH. *Health and Physical Education Score Card* No. 1, for Elementary Schools; No. 2, for Junior and Senior High Schools. Los Angeles: University of Southern California Press, 1938.
12. LARSON, LEONARD; FIELDS, MOREY; and GABRIELSEN, MILTON. *Problems in Health, Physical and Recreation Education.* New York: Prentice-Hall, 1953.
13. PHILLIPS, MARJORIE. "Problems of Questionnaire Investigation." *Research Quarterly* 12:528-37; October 1941.
14. SMITH, HENRY LESTER. *Educational Research, Principles and Practices.* Bloomington, Ind.: Educational Publications, 1944.

Attitude Scales

CHARLES C. COWELL

Many years ago Joubert, a French moralist, stated that "the direction of the mind is more important than its progress." Attitudes reflect the *readiness* of the organism to respond in certain specific ways when the proper situation arises. The fact that the organism is oriented and "triggered" to respond in a certain manner has resulted in speaking of attitudes as "*mental sets*" which exert a selective function, often without the aid of conscious consideration.

NATURE AND IMPORTANCE OF ATTITUDES

Educationally, we are interested in changing the behavior of students in desirable directions. We want them to develop desirable personal and social attitudes, e.g., attitudes toward maintenance of good health and the prevention of ill health, toward wholesome recreation, democratic ideals, and social improvement. When teachers and others build *readiness* in pupils to behave in these specific ways, for example, they are building *attitudes*. Finally, out of a number of general attitudes with some intellectual elaboration comes one's *philosophy of life* which reflects one's attitudes toward many phenomena (1).

The late President Eliot of Harvard, emphasizing the aim of education as the development of right attitudes and interests, spoke of liberal education as a "state of mind." Even if we think of an attitude as a point of view, we know that one's point of view determines what one sees, whether one sees clearly or not, and whether one sees things in the right perspective or not. Finally, the late Lord Halifax's definition of education as "what remains after we have forgotten everything we learned in school" indicates that he felt that the mental attitudes, permanent interests, and habits of study and thought acquired in school are the significant things.

Good education is an emotional as well as an intellectual experience. When we build feelings for or against something, we are developing *attitudes* (26). An attitude is "an implicit response or predisposition to act toward or away from an individual or social value" (4:176). It is generally agreed that attitudes are learned and they are developed solely in situations which call forth the attitude. For example, let us consider the situation of a fairly unskilled fifth-grade boy who has just come to this school and community from another city and faces a situation as follows:

1. He comes to class and is berated by the physical education teacher for not having "official" gymnasium clothing.
2. He is assigned a tiny locker in an overcrowded and odorous locker room.
3. He finds that the combination on his lock will not work.
4. At his first appearance in class he is asked by the instructor to do a certain exercise but fails miserably, injures himself, the class laughs at him, and the instructor makes a sarcastic remark.

This sequence of experiences, and the fact that all seem pointed in the same direction and suggest failure and unpleasantness to this boy, is reasonable assurance of a negative attitude—a dislike or even hatred of physical education and people related to it. He would hardly build an enduring *state of readiness* favorable to physical education and people related to it. His attitudes toward physical education or physical recreation activities, toward his teacher, and even toward his classmates might well be extremely *negative* and therefore educationally undesirable.

Fortunately, we have experimental evidence that attitudes are not fixed and unchanging predispositions (3, 23). Attitudes do change under normal conditions, and when conditions are controlled the changes may be striking. It is quite possible that the negative attitude toward physical education of the boy just mentioned could be changed to a positive attitude with systematic endeavor.

MEASUREMENT OF ATTITUDES

The problem of attitude measurement is complicated by lack of certainty regarding what we purport to be measuring. Bonner (4:195) raises the question:

> "If an attitude is a *tendency* to act, a predisposition to respond positively or negatively to an object, a more or less enduring *state of readiness* to respond in a certain way, can it be measured? If we measure an individual's verbal statement of what he thinks about an issue, are we measuring his attitude?"

In many cases there is only slight relationship between attitudes expressed in a paper and pencil test and the behavior of the subjects. Behavior does not always conform to expressed attitudes. An individual may *say* one thing and *do* another (11).

Like all human phenomena, attitudes, which are essentially disguised tendencies, are highly complex. If one is aware of the pitfalls and dangers and realizes that resultant conclusions of attitude tests are only partial insights into the total personality, these instruments can be of considerable value.

Statements referring to attitudes may likewise be applied to many related concepts which are either synonymous or vaguely synonymous and are tested with similar techniques in the technical literature. Such concepts as interests, desires, tastes, motives, opinions, morale, appreciation, ideals, and personal and social distance would fall into this category.

TECHNIQUES

The two techniques of getting individuals to manifest their *covert* tendencies in some form of *overt* behavior are:

1. More common is the *opinion* method, using techniques to obtain from the subject a *verbal* report of his attitude by checking the extent to which he agrees with various statements, the extent to which he *values* this or that, or by expressing his *opinion* about a certain object or state of affairs. An opinion is a verbal expression of one's attitude.

Thurstone's technique, known as the Equal-Appearing Intervals Scale, consists of getting a collection of statements that represent sentiment of approval or disapproval toward a phenomenon. These are sorted by a number of judges into categories from "Very favorable" to "Very unfavorable." Any significant disagreement by the judges over an item is cause for its rejection. The final scale distributes items in fairly equal steps along the continuum of items. The subjects then check the items of which they approve, and their score is the median scale value of the items checked. The scale value of each subject is then a relative measure of the amount of approval or disapproval accorded the phenomenon. A few illustrative items used in the finished scale to measure attitudes toward the church are reproduced below (29:61):

Check (x) every statement below that expresses your sentiment toward the church. Interpret the statements in accordance with your own experience with churches.

		Scale value
( )	1. I think the teaching of the church is altogether too superficial to have much social significance	8.3
( )	2. I feel that church services give me inspiration and help me to live up to my best during the following week	1.7
( )	3. I believe in what the church teaches but with material reservation	4.5
( )	4. I do not receive any benefit from attending church services but I think it helps some people	5.7
( )	5. I believe in religion but I seldom go to church	5.4
( )	6. I regard the church as a static, crystallized institution and as such it is unwholesome and detrimental to society and the individual	10.5
( )	7. I believe church membership is almost essential to living life at its best	1.5
( )	8. I believe the church is fundamentally sound but some of its adherents have given it a bad name	3.9
( )	9. I think the church is a parasite on society	11.0

Another attitude scale is Likert's (19) which employs a larger number of items than the Thurstone scale and does not employ judges in selecting items. The subject's approval or disapproval of a given phenomenon such as physical education, educational method, or social philosophy would be indicated by the degree of his agreement or disagreement with each statement on a five-point scale, such as "Agree markedly, Agree, Undecided, Disagree, Disagree markedly." The subject's total score is the sum of the item values weighted by how well each item in the scale distinguishes those who agree from those who disagree with each statement. A number of studies indicated in the references illustrate adaptations of the Likert technique (10, 19, 20, 32).

One of the problems in attitude testing is to determine the degree to which all statements or items in the instrument are related to the same attitude. This refers to the internal consistency or internal validity of the instrument, i.e., the degree to which the series of items which are given a single total attitude "score" are really interrelated and can be accepted as having a single attitudinal meaning rather than a mixture of different kinds of attitudinal responses. Guttman (14) suggested a procedure to solve this problem.

2. More promising but less highly developed is the *interpretive* or *projective* method, using techniques designed to lead the subject to betray his attitudes by expressing them without awareness of the investigator's purpose or design. Rather than having his responses structured for him, as in a questionnaire scale, the projective technique permits the subject to structure his own responses. The subject's attitude is determined by scoring certain agreed-upon indicators which reveal the subject's attitudes (13, 24).

It is difficult to apply the concept of validity to most projective tests since they do not yield single total scores or sets of scores having the same significance for all individuals. Furthermore, highly trained specialists are needed for their interpretation.

Another approach to attitude identification and measurement has been made by interpretation of "self-concepts." The sociologists, Kuhn and McPartland (18), have proposed a Twenty Statements Test (known as TST) which may be scored in various ways. This is a free-response type of test in which the subject is asked to write 20 spontaneous answers to the question "Who am I?" These authors and others have employed the test with different age groups. They have found that attitudes are revealed through the nonconsensual statements, their intensity through a saliency score derived from rank and/or frequency of the nonconsensual items, and the center of interest by a social anchorage score. This technique has been employed in some physical education studies (5, 15).

ATTITUDES AS EDUCATIONAL OUTCOMES

Attitudes are potent influences in individual and social control as the days of Hitler, Stalin, and Mussolini will attest. By consciously directing the culture in which their people lived and by planned attack upon the culture of any people they wished to dominate, they threw the world into chaos by developing attitudes of fear, hate, and prejudice.

Attitudes toward health (8, 27), physical education (2, 16, 31), recreation (21), and safety are important general objectives of instruction. We are interested in changing the direction of attitudes from negative and neutral to positive, and we must continue to seek objective and reliable means of measuring this change in direction. Since most systematic study of attitudes has dealt with indices of attitudes based on opinions expressed or approved, their validity depends upon the genuineness and frankness of the individual's response.

Knowledge by itself is inert. It becomes dynamic through motive, purpose, and desire which give it direction. Intelligence and knowledge determine what we *can* do. Our attitudes determine what we *will* do. Further research in the nature of attitudes, and in their development and measurement, is imperative for the better educational achievement of individual and social improvement.

REFERENCES ON ATTITUDE SCALES

1. ANGYAL, ANDRAS. *Foundations of a Science of Personality.* New York: Commonwealth Fund, 1941. p. 165.
2. BELL, MARGARET, and WALTERS, C. ETTA. "Attitudes of Women at the University of Michigan Towards Physical Education." *Research Quarterly* 24:379-91; December 1953.
3. BLUMER, H. *Movies and Conduct.* New York: The Macmillan Co. 1933.
4. BONNER, HUBERT. *Social Psychology.* New York: American Book Co., 1953.
5. BOWMAN, MARY O. *The Relationship Between Student and Parent Attitudes and Skills of Fifth Grade Children.* Doctor's thesis. Iowa City: State University of Iowa, 1958.
6. BROER, MARION R., and HOLLAND, DOLLY. "Physical Education Interests and Needs of University of Washington Women in Service Classes." *Research Quarterly,* 25:387-97; December 1954.

7. Buros, O. K., editor. *The Fourth Mental Measurements Yearbook.* Highland Park, N. J.: Gryphon Press, 1953.
8. Byrd, O. E. *Health Attitude Scale.* Stanford, Calif.: Stanford University Press, 1941.
9. Carr, Martha G. "Relation Between Success in Physical Education and Selected Attitudes Expressed by High School Freshmen Girls." *Research Quarterly,* 16: 176-91; October 1945.
10. Cowell, C. C.; Daniels, A. S.; and Kenney, H. E. "Purposes in Physical Education as Evaluated by Participants, Physical Education Supervisors, and Educational Administrators." *Research Quarterly* 22:286-97; October 1951.
11. Dawley, Dorothy J.; Troyer, Maurice E.; and Shaw, John H. "Relationship Between Observed Behavior in Elementary School Physical Education and Test Responses." *Research Quarterly* 22:71-76; March 1951.
12. Edwards, A. L., and Kilpatrick, F. P. "A Technique for the Construction of Attitude Scales." *Journal of Applied Psychology* 32:374-83; 1948.
13. Fromme, A. "On the Use of Certain Qualitative Methods of Attitude Research: A Study of Opinions on the Methods of Preventing War." *Journal of Social Psychology* 13: 429-59; 1941.
14. Guttman, L. "A Basis for Scaling Qualitative Data." *American Sociological Review* 9:139-50; 1944.
15. Isenberger, Wilma E. "Self-Attitudes of Women Physical Education Major Students and of Women Physical Education Teachers." *Research Quarterly* 30: 44-53; March 1959.
16. Kappes, Eveline E. "Inventory To Determine Attitudes of College Women Toward Physical Education and Student Services of the Physical Education Department." *Research Quarterly* 25:429-38; December 1954.
17. Kilander, H. F. "What the Public Knows About Nutrition." *Research Quarterly* 15: 188-97; May 1944.
18. Kuhn, Manford H., and McPartland, Thomas. "An Empirical Investigation of Self Attitudes." *American Sociological Review* 19:68; February 1954.
19. Likert, R. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22:5-43; 1932.
20. McCue, Betty Foster. "Constructing an Instrument for Evaluating Attitudes Toward Intensive Competition in Team Games." *Research Quarterly* 24:205-209; May 1953.
21. Moore, Beverly Y. "The Attitudes of College Women Toward Physical Activity as a Means of Recreation." *Research Quarterly* 12:720-25; December 1941.
22. Nemson, E. "Specific Annoyances in Relation to Student Attitude in Physical Education Classes." *Research Quarterly* 20:336-47; October 1949.
23. Peterson, R. C., and Thurstone, L. L. *Motion Pictures and the Social Attitudes of Children.* New York: The Macmillan Co., 1933.
24. Proshansky, H. M. "A Projective Method for the Study of Attitudes." *Journal of Abnormal and Social Psychology* 38:393-5; 1943.
25. Remmers, H. H. *Introduction to Opinion and Attitude Measurement.* New York: Harper & Brothers, 1950.
26. Remmers, H. H., and Gage, N. L. *Educational Measurement and Evaluation.* Revised edition. New York: Harper & Brothers, 1955.
27. Southworth, W. H.; Latimer, J. V.; and Turner, C. E. "A Study of Health Practices, Knowledge, Attitudes, and Interests of Senior High School Pupils." *Research Quarterly* 15:118-36; May 1944.
28. Stalnaker, John M. "Attitudes Toward Intercollegiate Athletics." *School and Society* 37:499-504; 1933.
29. Thurstone, L. L., and Chave, Ernest J. *The Measurement of Attitude.* Chicago: University of Chicago Press, 1929.
30. Wang, Charles K. A. "Suggested Criteria for Writing Attitude Statements." *Journal of Social Psychology* 3:367-73; 1932.

31. WEAR, CARLOS L. "Construction of Equivalent Forms of an Attitude Scale." *Research Quarterly* 26:1 3-19; March 1955.
32. WEAR, CARLOS L. "The E luation of Attitudes Toward Physical Education as an Activity Course." *Research Quarterly* 22:114-26; March 1951.
33. ZIMMERMAN, HELEN M. "Physical Education Experiences and Interests of College Women." *Research Quarterly* 25:109-18; March 1954.

Sociometric Techniques

CHARLES C. COWELL

The researcher reasons with data about the solution of a problem. The soundness of the solution depends upon the validity of the facts or principles upon which the inferences are based—upon the validity of the data.

SOCIAL RELATIONS

Physical education teachers and recreation leaders are, in a sense, "development supervisors." They are interested in all aspects of development—orderly progress toward maturity of boys and girls. Therefore, the degree of harmony of the individual with his social group and his social growth from year to year become an educational concern. Here the playground, athletic field, swimming pool, and gymnasium provide the observant teacher with valid data indicative of the pupil's social conduct, his degree of social feeling, and acceptance by his peers. The pupils' social roles and their relationships with their peers play a significant part in the process of socialization and personality development.

This section deals with *sociometrics,* which has to do with the study of the patterned relationships between members of groups. Children and youth struggle for "belonging" and group status. The roles they learn to play in achieving status are important sociologically, psychologically, and educationally. Teachers, therefore, must help provide the resources in the form of activities and experiences which lead to effective goal satisfactions—the winning of belonging and recognition. *Sociometrics* as a technique quantifies the degree to which the pupil is winning belonging and

3

prestige status by playing roles effectively in the peer-group structure.

During middle childhood and adolescence, play, physical education, recreation, and certain health activities evoke strong interest. Success in these activities provides powerful goal satisfactions and varied degrees of prestige. The individual's place in the group is largely determined by the various roles he plays, the degree of success in these roles, and the prestige value of the various roles.

All teachers are familiar with the "fringer," the "dub," the "isolate," and the "nonbelonger," but not all are familiar with the self-regard feeling, attitude, and concepts that such youths have of themselves and the part these concepts play in social and emotional adjustment. Lack of physical strength, lack of game skills, poor physical development, poor social relationships, feelings of inferiority, insecurity, inadequacy, and a distorted self-concept seem to be interrelated in middle childhood and adolescence.

The peer-group activities—in clubs, in the gymnasium, or on the playing fields—are important sources of learning.

> Through these activities each individual is disciplined by group processes to subordinate personal desires to the success of the group, to accept group customs and codes, needs and roles, to achieve personal success and status through successful group activities, to respect the rights of others, and to promote the purposes of the group as a whole. To a considerable extent it is through peer-group activities that leadership capacities are developed, the concept of teamwork is established, and a sense of personal adequacy based on sure belonging engendered. (22:278)

Surely the acquisition of these attributes is just as important for success in life as is the acquisition of the standard curriculum learnings usually called "subject matter."

Since group life is of such importance in the development of children and youth, some consideration of the structure and dynamics of social groups as met in health, physical education, and recreation activities will be given in this section.

DEFINITIONS

Sociometry represents techniques for presenting rather simply and graphically the structure of interpersonal relations within a given group and for quantitatively studying its internal organization.

A *sociometric test* is a social research instrument for diagnosing, understanding, and evaluating the structure of a group by noting the social cleavages involving acceptance and rejection, and for locating the leaders as well as the isolates as these appear in the analysis of the choice process.

A *sociogram* is a graphic diagram of a social structure illustrating the pattern of relationships between members of a group. It indicates graphically who accepts whom, who is rejected by whom, and the nature of social cleavages, such as cliques.

TYPES OF SOCIOMETRIC TESTS

The spontaneous choice of one's associates is the index used in the sociometric test. The choices refer to those persons with whom the individual would like to participate in a given situation—on a team or committee or working unit.

Types of tests especially adaptable to physical education situations are described below.

1. *The Acquaintance Volume Test* indicates the "social expansiveness" of an individual within a given time period. It shows how well a student gets acquainted within a semester. As suggested by Todd, "On the first day the class meets, each member is asked to write the first and last names of those he knows in the group. At the end of the unit or term, the test is repeated, and by simple arithmetical differences it is readily apparent just how many new friends each individual has made." (29) Skubic (27) applied this type of test successfully in studying the differences in acquaintance volume in different activity classes such as volleyball, swimming, and modern dance.

2. *The Functional Choice Test*, as described by Todd, "is a means of finding out who *wants* to be with whom—not who *is* with whom. Members of the group are given opportunity to choose or reject others on some specific basis in which existing friendship bonds, and a promise to take them into consideration in regrouping the class, are motivational factors." (29)

3. *The Cowell Personal Distance Ballot* (5) asks each class or group member to indicate the personal distance at which he is willing to accept every other member. With this instrument, it is possible to note the general attitude of the class as a whole toward each individual as well as to determine quantitatively the degree of acceptance of each individual in the group by every other individual in the group. It is thus possible to know the personal distance at which each group member would hold every other.

The individual's personal-distance score indicates the group's generalized attitude of acceptance toward each individual since—unlike the true sociometric test—it provides no specific frame of reference for choosing, e.g., choosing committee chairmen, team members, or roommates. Each individual is checked on a seven-point scale of personal distance, the first being "Acceptance into my family as a brother (or sister)," and the seventh being "Into my city" with various progressive distances represented in between. This instrument is an adaptation of ideas used by Bogardus (1) in his "Test of Social Distance" and has been found simple, valid, and reliable (5).

ADMINISTRATION OF SOCIOMETRIC TESTS

Todd (29) illustrates the use of the *Functional Choice Test* by casually asking the pupils to write on 3 x 5 file cards their first three preferences for squadmates for the ensuing activity. "Promise the class that their choices will be kept confidential and that you will guarantee to place each pupil on a squad with at least one of his chosen friends—more if you can. Suggest that if there is anyone with whom the pupil would prefer *not* to play, that the name (or names) of that person be indicated at the bottom of the card." This is a five-minute task. Data are tabulated in several ways:

1. A *matrix chart* (4) shows each individual's preferences. (See Figure I.) Students' cards are alphabetized and numbered, and the names listed down and across the chart. Degrees of popularity and unpopularity are readily noted and rejections emphasized by red pencil.

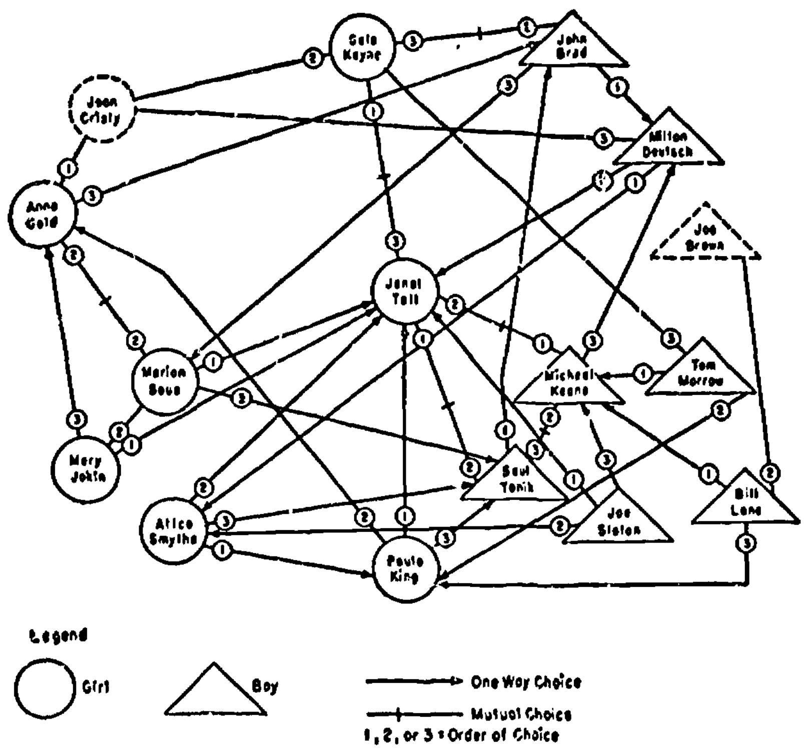
Chooser → / Chosen ↓		Ann 1	Betty 2	Chloe 3	Donna 4	Eve 5	Fran 6	Joy 7	Kate 8	ISI*
1	Ann	—		R		1		3	2	
2	Betty		—		2	1	3			
3	Chloe			—	2		3	1		
4	Donna		1		—	3	2			
5	Eve	1	3			—		R	2	
6	Fran	1			3		—	2		
7	Joy		1		2	3		—		
8	Kate	1				2		3	—	
Total Choices		3	3	0	4	5	3	4	2	GCS**
Total Rejections		0	0	1	0	0	0	1	0	

Key
1, 2, 3 = order of choice
R = rejection
* = individual status index
** = group cohesion score

FIGURE I. Matrix chart.

2. *A sociogram* is a graphic diagram indicating the relationships between individuals in a group. One may note the cliques, gangs, pairs, and the extremely popular, as well as the "rejectees" or "isolates" (24, 11).

The usual procedure is to place the most popular near the center, the boys on one side, girls on the other as indicated in Figure II.



NOTE: For an absent boy or girl, use the respective symbol dashed, leaving any choice line open-ended (see Joe Brown above).

If rejections are obtained, the choice line may be made in dashes or in a different color.

Whenever a direct line from chooser to chosen cannot be drawn without going through the symbol for another individual, the line should be drawn with an elbow, as in the case of Bill Lane to Paula King.

FIGURE II. A filled-in sociogram, presenting the choice patterns graphically. Blank forms with empty circles and triangles may be mimeographed so that the teacher may fill in the names and draw in the choice lines after the test has been given. (Quoted by permission from Jennings, 11: 22.)

3. The *Individual Status Index* (4, 29) is a simple formula for assigning to each student a numerical symbol of his popularity with his peers. Retests during and at the end of the term will then show any changes in the pupil's status. The formula follows:

$$\text{Individual Status Index} = \frac{\text{Total Choices minus Total Rejections}}{\text{Number in class minus one}}$$

or

$$\text{ISI} = \frac{\text{TC} - \text{TR}}{\text{N} - 1}$$

For example, if, in a class of 50, Mary receives 5 choices and 1 rejection, her ISI is .08.

4. *Group Cohesion Scores* indicate the degree of social integration or "we feeling" within a group. Retests will show increases or decreases in group solidarity. This may serve as a check on subjective observation. The formula follows:

$$\text{GCS} = \frac{\text{TC} - \text{TR}}{\text{N (N} - 1)}$$

For example, if in a class of 50 there is a total of 110 choices and 10 rejections, then the GCS = +.02.

INTERPRETATIONS, APPLICATION, AND FOLLOW-UP

Knowledge of the social structure of a group is important but, perhaps, more important is some understanding of the personality dynamics which sociometric devices suggest. We must try to analyze the choice process going on within people when they select some individuals and reject others. As teachers, we have some control over social participation of children and youth and the range of their social contacts. Since their social maturity is dependent on their social interaction with others, we should *plan* significant ways of preventing social cleavages and encourage social integration in joint action.

Diagnosis should always precede prescription. Sociometric techniques enable us to diagnose and evaluate the social structure of groups, help us to locate the rejected pupils, and regroup classes so that more children are given a sense of belonging in a more productive and harmonious group atmosphere.

REFERENCES ON SOCIOMETRIC TECHNIQUES

1. Bogardus, E. S. "A Social Distance Scale." *Sociology and Sociological Research* 17:265-71; 1933.
2. Bonney, Merl E. "Sociometric Study of Agreement Between Teacher Judgements and Student Choices." *Sociometry* 10:133-46; 1947.

3. COOK, LLOYD ALLEN. "An Experimental Sociographic Study of a Stratified 10th Grade Class." *American Sociological Review* 10:250-61; 1945.
4. COOK, LLOYD A. "Manual for Obtaining, Analyzing, and Diagramming Sociometric Data." Detroit: College Study in Intergroup Relation, 1950. (Mimeographed)
5. COWELL, CHARLES C. "Validating an Index of Social Adjustment for High School Use." *Research Quarterly* 29:7-18; March 1958.
6. EDWARDS, A. L., and KILPATRICK, F. P. "Scale Analysis and the Measurement of Social Attitudes." *Psychometrika* 13:99-114; 1948.
7. FULTON, RUTH, and PRANCE, ELIZABETH. "Motor Learning of Highly Chosen and Unchosen Teammates and Friends." *Research Quarterly* 21:126; May 1950.
8. GRONLUND, NORMAN E., and WHITNEY, ALGARD P. "Relation Between Pupil's Social Acceptability in the Classroom, in the School, and in the Neighborhood." *The School Review* 60:267-71; 1956.
9. GUTTMAN, L. "A Basis for Scaling Qualitative Data." *American Sociological Review* 91:139-50; 1944.
10. JENNINGS, HELEN HALL. "Sociometric Grouping in Relation to Child Development." *Fostering Mental Health in our Schools.* Washington, D. C.: Association for Supervision and Curriculum Development, a department of the National Education Association, 1950. Chapter 13.
11. JENNINGS, HELEN H. *Sociometry in Group Relations.* Washington, D. C.: American Council on Education. 1948.
12. KOZMAN, HILDA, editor. "Newer techniques of promise." *Group Process in Physical Education.* New York: Harper & Brothers, 1951. 405 p. Chapter 7.
13. KUHLEN, RAYMOND G., and BRETSCH, HOWARD S. "Sociometric Status and Personal Problems of Adolescents." *Sociometry* 10:122-32; 1947.
14. KUHLEN, R. G., and COLLESTER, E. G. "Sociometric Status of Sixth and Ninth Graders Who Fail to Finish High School." *Educational and Psychological Measurements* 12:632-37; 1952.
15. MCGRAW, L. W., and TOLBERT, J. W. "Sociometric Status and Athletic Ability of Junior High School Boys." *Research Quarterly* 24:72-80; March 1953.
16. MARKS, J. B. "Interests, Leadership and Sociometric Status Among Adolescents." *Sociometry* 17:340-49; 1954.
17. MORENO, J. L. *Who Shall Survive?* New York: Beacon House, 1934.
18. MOUTON, JANE SRYGLEY; BLAKE, ROBERT R.; and FREICHTER, BENJAMIN. "The Validity of Sociometric Responses." *Sociometry* 18:181-206; 1955.
19. NEWCOMB, T. M. "Social Attitudes and Their Measurements." In MURPHY, G.; MURPHY, LOIS B.; and NEWCOMB, T. M. *Experimental Social Psychology.* Revised edition. New York: Harper & Brothers, 1937.
20. NORTHWAY, MARY L. *A Primer of Sociometry.* Toronto: University of Toronto Press, 1952.
21. PHILLIPS, BEEMAN N., and DEVAULT, M. VERE. "Relation of Positive and Negative Sociometric Valuations to Social and Personal Adjustment of School Children." *Journal of Applied Psychology* 39:409-12; 1955.
22. PRESCOTT, DANIEL A. *The Child in the Educative Process.* New York: McGraw-Hill Book Co., 1957.
23. PROCTER, C. H., and LOOMIS, C. P. "Analysis of Sociometric Data." in JAHODA, MARIE; DEUTSCH, M.; and COOK, S. W., editors. *Research Methods in Social Relations,* Part II: Selected Techniques. New York: Dryden Press, 1952.
24. REMMERS, H. H., and GAGE, N. L. *Educational Measurement and Evaluation.* Revised edition. New York: Harper & Brothers, 1955. p. 335.
25. SATTERLEE, ROBERT L. "Sociometric Analysis and Personality Adjustment." *California Journal of Educational Research* 6:181-84; 1955.
26. SHOOBS, NAHUM. "Sociometry in the Classroom." *Sociometry* 10:154-64; 1947.
27. SKUBIC, ELVERA. "A Study of Acquaintance and Social Status in Physical Education Classes." *Research Quarterly* 20:80-87; March 1949.

28. THORPE, J. G. "An Investigation into Some Correlates of Sociometric Status Within School Classes." *Sociometry* 18:49-61; 1955.
29. TODD, FRANCES. "Sociometry in Physical Education." *Journal of Health, Physical Education, and Recreation.* 24:5:23, 36; May 1953.
30. ULLMAN, C. A. *Identification of Maladjusted School Children.* U. S. Office of Education, Federal Security Agency, Monograph No. 7. Washington, D. C.: Superintendent of Documents, Government Printing Office, 1952.
31. WALTERS, C. ETTA. "A Sociometric Study of Motivated and Non-Motivated Bowling Groups." *Research Quarterly* 26:107-12; March 1955.
32. YOUNG, L. L. "Sociometric and Related Techniques for Appraising Social Status in an Elementary School." *Sociometry* 10: 168-77; 1947.

Photography

ALFRED W. HUBBARD

Prints, slides, strip film, loop film, and motion pictures reveal so much more realistic detail than unaided observation that photography seems an ideal research tool for studying human movement and sport skill. The ability of cameras to record visible, transient phenomena, to enlarge or reduce spatial relations, and to slow down or speed up action has made them highly useful in research, in addition to providing realistic illustrative material, instructional aids, and nostalgic records of people and places. The wealth of photographic material about sport and sporting events seems like a gold mine of material for scientific analysis, but this illusion generally evaporates when measurement starts.

Cameras are basically of two types—still and motion picture. But motion can be recorded with still cameras, either intentionally or accidentally, and motion pictures consist of a sequence of still pictures (frames). Both types provide a permanent record for detailed analysis and facilitate interpretation of human movement in terms of basic mechanical (Newtonian) principles. As long as the analysis will be based on inferences concerning the operation of these mechanical principles, any clear photographic record is suitable. But scientific analysis involves measurement, and measurement of movement requires accurate recording of spatial and temporal relations. These two basic, and hidden, limitations exclude practically all sports movies from scientific analysis, unless the material was taken especially for the purpose of providing meaningful measurements.

Movement is displacement with respect to time; displacement is a change in spatial relations; and time is a convenient framework of reference. Valid measurements of movement thus depend on accurate recording of spatial and temporal relations. Unless the photographer makes special arrangements to record comparable distances at uniform intervals in standard units of measurement, the analyst of human movement is restricted to vague estimates and unsupported inferences. These limitations are hidden, or often unrecognized by the novice. The camera operates basically like the human eye, so photographic material that appears highly realistic may actually provide no valid basis for measurement. In using photography as a research tool for studying human movement, the first problem is to record comparable spatial relations and the second problem is adequate timing of the changes in spatial relations. Fortunately, both problems can be solved essentially by rather careful but simple, preliminary preparation which avoids the omission of essential elements in the photographic record and thus makes the pictures suitable for scientific analysis.

SPATIAL RELATIONS

The camera, like the human eye, reduces three-dimensional space to two dimensions, or a plane surface. Near objects appear large, and far objects small. Since the visual world is full of illusions caused by perspective, the real size of things must be judged ordinarily on some basis other than apparent size. Thus, the distant mountain appears many times higher than the tree in the foreground that apparently looms over it. People learn to disregard these illusions so consistently that any clear photograph seems excellent, until measurements are attempted. Then it is found, generally after considerable confusion, that only those spatial relations occurring in one plane perpendicular to the axis of the camera are directly comparable. Any appreciable movement or displacement in the third (to-from) dimension is enlarged or diminished. Measurements involving varying amounts of "to-from" distortion are not comparable. This makes photography for measurement purposes a highly specialized technique—for which the accumulated "wealth" of photographic material concerning sport is fool's gold. In other words, if measurements are to be made, the camera must be properly placed.

Fortunately, much human movement of the body or its parts occurs in one plane, or essentially one plane. Thus, to make measurements and comparisons based on photographic records, the *major movement plane* must be determined and the camera set perpendicular to and approximately at the center of this plane. Then movement in this plane can be measured and compared. Any appreciable movement in the third (to-from) dimension cannot be measured accurately or compared (except with a second, synchronized camera and very complicated procedures). However, to-from components of movement can be observed and inferences drawn, as is done in visual observation.

More specifically, directly comparable measurements can be made only from pictures taken perpendicular to the plane of the movement and at a *fixed distance*. In actual practice, the distance from camera to subject and enlargement on projection may both vary. Therefore, a *scale object* should be included in the photographic field. A crossbar with alternate black and white stripes one foot wide, a six-foot or two-meter distance marked with chalk or tape on apparatus, or anything of known size in the plane of the movement can be used as a scale object. Knowing the actual size of the scale object and measuring the apparent size on the print or projection makes possible conversion of measurements to actual units. Actual distance : apparent distance :: actual scale length : apparent scale length—i.e., actual distance equals apparent distance times the ratio of actual to apparent length of the scale object. One conversion factor will suffice if uniform distance from camera to subject and uniform enlargement are used, but a scale object must be included to ensure uniformity.

TIMING

Stretching the time base of movies (taking pictures rapidly and projecting them slowly) aids observation. Slowing movement to a half or a quarter of its normal speed permits more detailed observation without losing the flow of movement. Equipment is available for much greater "stretching," but with this, slow movements stand still and fast movements lose their characteristic flow. In human movements consisting of slow wind-ups and fast strokes, the desirable amount of stretching is a compromise between getting interminable wind-ups and too few frames of the fast stroke, or between wasting film and not getting enough frames of the fast portion for adequate study. The only solution is to run the camera fast enough for the specific job, and forget about the cost. The real problem, if measurements leading to computations of velocity and velocity changes are to be made, is determination of the interval between frames and the frames per second—i.e., timing the exposures.

Movie cameras with internal timers recording on film are available, but expensive. Most cameras suitable for cinematographic analysis of sport skills are spring driven and governor controlled, with a variable speed selector in terms of frames per second. If kept wound and warm, they accelerate rapidly and maintain a reasonably constant speed. However, the speed selector indicates only the approximate number of frames per second and, for various reasons, is not sufficiently accurate for precise work.

If the speed selector is not changed from sequence to sequence, accidentally or intentionally, timing can be obtained by photographing a falling object (anything solid, reasonably heavy, and round) and using the formula $gt^2/2$. The acceleration of gravity (g) is 32.2 feet per second per second, so $g/2$ is 16.1 (feet/second^2).[1] A fairly accurate method, described by Cureton (4), consists of photographing an object dropped 8 feet (which takes .705 seconds using 16.1 as $g/2$) and dividing the time by the number of frames from release to contact to get the interval between frames, or time/frame. The reciprocal (one divided by this interval) is frames per second. This system has two appreciable sources of error—release and contact occur generally between frames, and the exact frame in which release occurs is difficult to determine because the object moves very slowly immediately after release.

A more accurate method consists of dropping an object beside a scale so that its fall can be measured in two frames—one near

[1]The acceleration of gravity (g) increases uniformly the rate at which bodies fall by 32.2 feet per second each second, or 32.2 feet per second per second (ft./sec.^{2}). A body at rest has an initial velocity (V_o) of zero feet per second. A body falling from rest has a terminal velocity (V_t) of g multiplied by the number of seconds (t) that it falls, or gt. The distance (s) that it falls is the average velocity $[(V_o+V_t)/2]$ multiplied by the time (t in seconds). Substituting 0 for V_o and gt for V_t gives an average velocity of $gt/2$. Multiplying average velocity by time to get distance gives $s = gt^2/2 = (16.1)t^2$.

the start and the other near the ground. The time at which an object dropped from a known height would reach the upper and lower points can be calculated from the formula, $s=gt^2/2$. The difference is the elapsed time, and this divided by the elapsed frames is the time interval between frames. The reciprocal is frames per second.[2] These methods are reasonably accurate, *provided* factors affecting camera speed have not changed.

An economical alternative to internal timing, and a simpler and more reliable method than photographing dropped weights, is to include a timing device in the photographic field. The external timer may be a high speed electrical counter actuated by hundredths or thousandths of a second pulses, but a large synchronous clock with a hand making one circuit per second and with divisions of hundredths of a second is simpler. Either must be closer to the camera than the subject to give a clear image, and this is possible only with sufficient illumination to use a relatively small aperture which gives sufficient depth of focus. An example of a synchronous clock appears in the top row of frames in Figure III. The outer case was removed to avoid glare from the glass. Black tape darts were affixed at each .05 second mark and also on the sweep second hand.

The clock is left running during the filming, but the time of each frame need not be read. Time is read near the beginning and near the end of an exposure in seconds and estimated in hundredths of a second. With clear .05 second marks, the maximum error should be less than .01 second. The error in the two readings tends to average zero, but if it does not, distribution over the number of frames makes it negligible. Dividing the elapsed time by the number of frames intervening between those in which the time was read (counting the first as "Frame O") gives time between frames (computed to at least ten thousandths of a second) and the reciprocal is frames per second. (Given 88 frames and 1.89

[2]For example: a weight dropped from 8 feet was found to be 7 feet 4½ inches above the ground in one frame and 10¼ inches above the ground 21 frames later. The first drop would be 7.5 inches, or .625 feet, and the second would be 7 feet 1¾ inches, or 7.146 feet. With $s=gt^2/2$, $t=\sqrt{2s/g}$, and $g=32.2$ ft./sec.^{2}: for .625 feet, $t=\sqrt{1.250/32.2}=\sqrt{.038820}=.197$ sec., and using 7.146 feet, $t=\sqrt{14.292/32.2}=\sqrt{.443851}=.666$ seconds. The difference (.666—.197) indicates that .469 seconds elapsed during the 21 frames, so the interval between frames (.469/21) is .022333 seconds. The reciprocal (1/.022333) gives 44.8 frames per second.

seconds elapsed time: 1.89 sec/88 frames gives .0215 sec. between frames, and 1 sec/.0215 sec. gives 46.5 frames per second.) Repeating this procedure for several exposures will indicate whether the camera speed was reasonably uniform, or whether a time base must be calculated for each exposure.

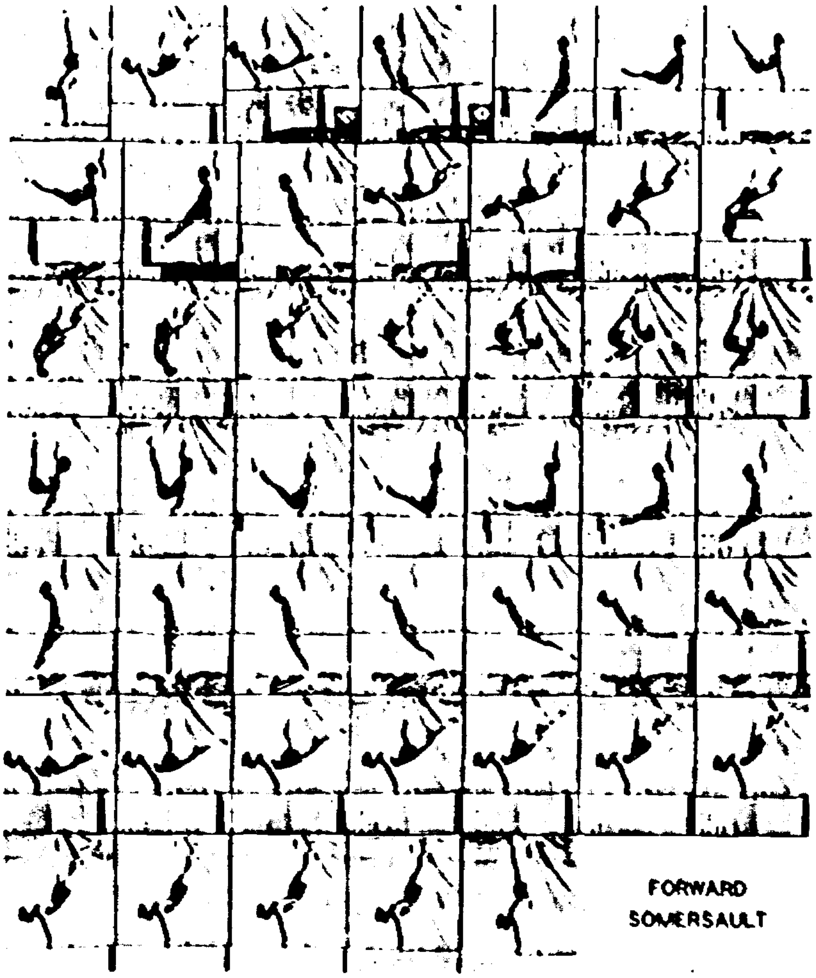


FIGURE III. Motion picture frames showing movement sequence in performing a forward somersault. From Sullivan (47).

MOTION WITH A STILL CAMERA

The prevalence of movies has fostered the notion that motion picture equipment is necessary for recording movement. Actually, methods for recording movement with a still camera, which were developed by Marey (38, 39) about 70 years ago and other methods introduced 30 years ago by Edgerton (10), are still useful. The simplest method, developed by Marey and still used in motion study, is to affix a small light source to the moving limb or object and have the cursory light trace the pattern of movement (30, 43). Speed of movement can be inferred, or even measured, from the width of the tracing, which narrows as the light moves faster and widens as it slows down. A refinement of this is to interrupt the moving light at regular intervals and to affix lights at several joints to show the relation of segments in the movement pattern. Another method for showing successive segmental movement was developed by Marey (38:61). He interrupted the illumination of the subject, who wore black underwear with white strips representing the bodily segments and moved before a dark background. This gave a series of "stick figures"—similar to those which appeared recently in *Life* (33).

Edgerton (10) generally used a motion picture camera in his movement studies, but he also used superimposed prints taken with a still camera. His special contribution was the development, with Gemerhausen, of a high-intensity, high-frequency light source (stroboscopic light) for high speed chronophotography. A patent for the idea of instantaneous pictures (using lightning) was issued in England before 1850, but Edgerton made this so practical that press photographers now use electronic photoflash equipment in place of flash bulbs. The equipment is similar, but flashing the light from 500 to 50,000 times a second is quite different from once in five seconds. However, recording movement with a still camera has one distinct advantage—the evidence is on one negative or print, not on 25 or 100 which must be collated. The analytical methods are similar to those used with motion pictures.

CINEMATOGRAPHIC ANALYSIS

After one takes the trouble to produce motion (or still) pictures with measurable spatial and temporal relations, the obvious thing is to start measuring and making comparisons. For maximum detail and ease in making reproductions, 35-millimeter negative movie film is generally used. If negative film is used, it should be inspected for quality and, if the negative is good, a positive contact print should be made. To attempt to measure from projected negatives leads to insanity. Various systems and equipment have been devised for measuring from projected positives frame by frame (20), but the simplest method is to use a microfilm reader, which is found in most libraries. These measurements have to be converted to standard units by using a conversion factor based on the real and apparent length of the scale object. By adjusting the size of the projected image and choosing suitable graph paper or measuring units, the conversion factor can be made a simple, rather than a complicated, multiplier, and this facilitates conversion. The scale object also serves to align frames. A motion picture camera, even though securely mounted on a tripod, vibrates enough to catch successive frames in slightly different positions. Consequently, a reference line, two points, or the outline of the scale object must be established and the projected image must be adjusted by moving the projector or the paper to make this coincide in successive frames.

However, before one becomes engrossed in the almost endless job of measuring, it is necessary to stop and consider. The purpose of measurement is to test something. Newton's laws of motion have been well tested and accepted, at least for our purposes. Some measuring will be necessary for calibration—or for testing the consistency and reliability of the photographic record. But the real purpose of cinematographic analysis of sport skills is to see how these laws apply, to show others how they apply, and eventually to improve our understanding of the physical problems in sport skills and our instruction of them. To do this, often good and bad performers, the skilled and the unskilled, or subjects obviously differing in skill are photographed. Then the problem is to find how the individuals or groups differ at various levels of ability. An old, ingenious, and still acceptable shortcut exists for helping to see these differences, which can later be verified by meaningful measurements.

Stick Figures. The shortcut is stick figures, which reveal a great deal after you learn to make and interpret them. Choose some reference points on the subject, such as the tip of the toes, the outer malleolus, the center of the hip, knee, and shoulder, the tragus of the ear, the center of the elbow and wrist, and the tip of the middle finger. Start from some definable point in the movement sequence, plot the reference points successively frame by frame (each time aligning the picture), and connect the corresponding marks with straight lines, except for the curved hip-shoulder line, as in Figure IV (47). These stick figures correspond to frames 61 through 74 in Figure III. The resultant graph is much less cluttered than a series of body outlines and shows considerably more. The numbers beside the head (tragus) and toes indicate the successive frames after release from the parallel bars, through the somersault, to a catch. The successive body positions and movements of parts now stand out clearly in their essential relationships.

This particular sequence (Figure IV) shows something that is often discussed in kinesiology, often demonstrated with waving arms, but seldom seen in print—translation of energy from one body segment to another. Note the spacing between Frames 1 and 2, and between 2 and 3 for the head and toe. Both are traveling at uniform velocity (equal space in equal time) during this phase. But the spacing between 3 and 4 and between 4 and 5 increases (acceleration) and then becomes essentially constant from Frame 5 through Frame 8. But the toe shows deceleration (less space) between 3 and 4 and between 4 and 5, and then uniform, lower velocity from 5 to 8. Then the head slows down and the feet speed up to about their original velocity. The whole body presumably rotates uniformly, but speeding up one part produces an equal and opposite reaction (slowing down) of another part, which illustrates the translation of energy.

Besides being simple and easy to construct, stick figures reveal relations and velocity changes of bodily parts that may easily remain hidden in measurements. If the successive distances the parts moved had been meticulously measured, added, and averaged, a set of (rather meaningless) average velocities would have resulted, and the fact that significant velocity changes occurred would never have been discovered. In the study from which these figures were taken (47), five subjects varying in ability to perform the front somersault on the parallel bars were photographed and stick figures were made for each. Comparison between subjects showed the differences upon which successful execution of

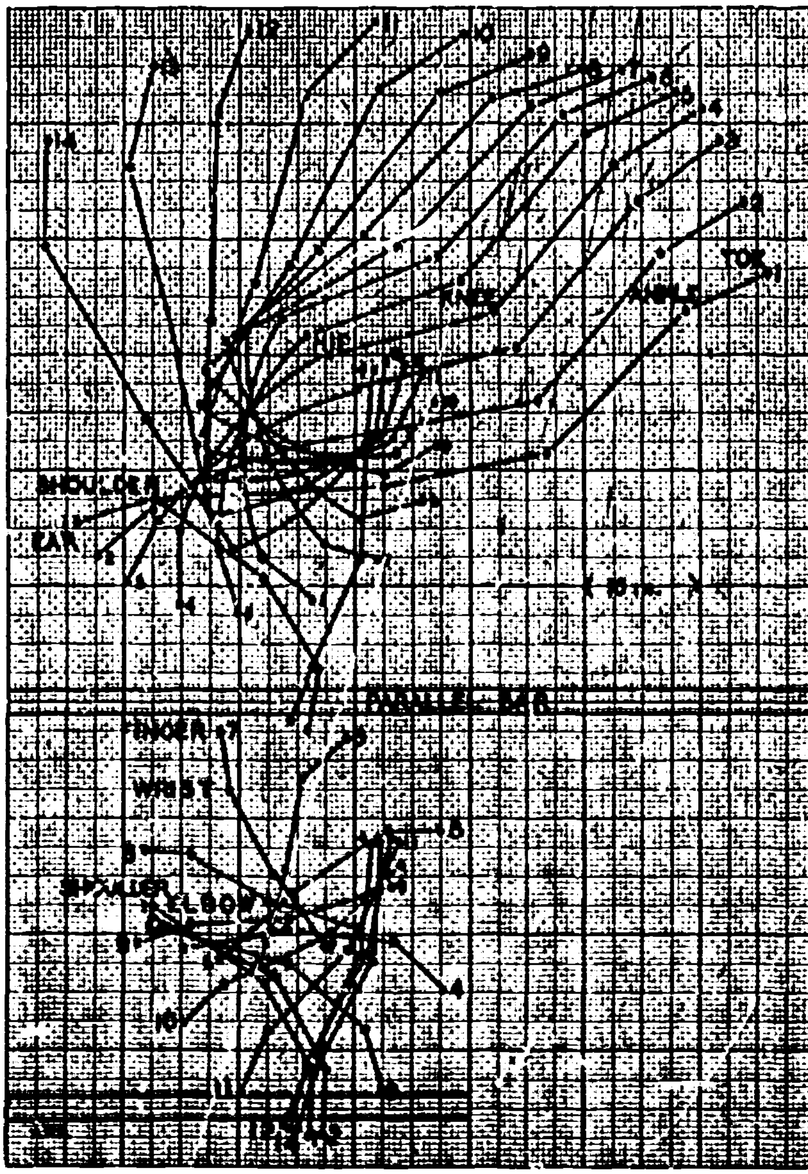


FIGURE IV. Stick figures showing successive positions of bodily segments from release (1) to regrasp (15). From Sullivan (47).

the stunt depended. From these, key differences and coaching points were deduced, and the range of skill indicated which individual differences were significant for successful completion of the stunt. Stick figures are by no means new, but they are simple, easily made, and revealing. It might be noticed that the arm action was drawn separately (lower part of Figure IV) to avoid the confusion of overlapping parts.

Center of Gravity. Forces producing the translation, rotation, or projection of bodies can be treated as though they act on, or in relation to, the center of gravity. In a body with movable parts, such as the human body, the center of gravity moves with respect to a fixed point in the body as the parts change their relationship. The center of gravity is the point about which the separate masses times their distances sum to zero. Therefore, as the parts move, the center of gravity moves. However, in projection (bodies flying through space), the center of gravity follows a parabolic path from loss of contact with a solid base to return, i.e., from release to contact.

Location of the center of gravity is a rather complicated procedure if one uses the classic methods of Braune and Fisher (2, 28). Applying their methods to 20 successive frames for five or six subjects performing some stunt might take two to four months and would involve very complex computations. A simpler method is to find out how the center of gravity moves in the human body with movement of the parts and to estimate its position in successive frames (after aligning each frame). The results of such an analysis are very interesting, but too often something other than a parabola results. Since the laws of mechanics indicate that the path is parabolic, the estimated positions must be in error. The system is not recommended.

A system for locating the position of the center of gravity in successive frames was reported by Groves (21). This system is much simpler and more direct than that of Braune and Fisher. The equipment consists of a 5/8 inch plywood panel, 6 feet square, painted black, with two white center lines dividing the panel into four quadrants. The panel is suspended at each corner from an accurately calibrated (Chatillon) scale. From movies of the stunt and prints of the successive positions, the subject is arranged on the board in positions corresponding to those in the stunt. For each position, the net weight on each scale is read and marked with chalk in the corresponding quadrant. Then the subject and recorded weights are photographed from directly overhead. The position of the center of gravity, with respect to the center of the board and the subject, is computed from the combined weights. The computation is well described in the source. It appears complicated but is not really difficult. The next step is to transfer the centers of gravity as determined to the successive frames of the stunt. Barring computational errors, this method locates accurately the subject's center of gravity for the position *on the board*. Consequently, the accuracy of these locations depends on the accuracy with which the position of the subject's bodily parts on the board duplicates their position in space for that frame, which is partly a matter of judgment.

A simpler, less laborious, and probably more accurate determination of successive positions of the center of gravity of the body in flight can be obtained directly from two frames. Since the procedure has not been published previously, the basic principles and necessary steps must be presented in some detail. After one labors through it once, it can be done relatively quickly. Basically, the path of the center of gravity of a projected body is a parabola. This parabola has two components, horizontal and vertical, and is symmetrical about the vertical axis. The horizontal component of motion is uniform; equal distances are traversed horizontally in equal units of time (frame by frame). Vertical travel is also uniform, but not in the sense of the distances being equal, since it depends, in essence, only on the acceleration of gravity ($gt^2/2$). By determining with reasonable accuracy the location of the center of gravity at release (loss of contact with a solid base) and contact (regaining contact), the parabola of flight can be constructed accurately, or accurately enough for graphing.

The first step is to locate the frames in which release and contact occur. Since both probably occurred between frames, the frame chosen for release should be that immediately preceding the one in which support was lost clearly and that chosen for contact, the frame in which support was first regained clearly. A sheet of graph paper should be aligned with a known vertical or horizontal in the projected image, the scale object should be indicated, and then the body outline should be traced at release. After realignment of the paper, if necessary, the body outline should be traced at contact. This gives two, perhaps partly overlapping, body outlines—one for release and the other for contact.

The second step is to determine the center of gravity of the body in each position. The center of gravity is the point about which the moments (body masses times their distances) are equally distributed—sum to zero. Using a 6- or 8-inch, 360° transparent protractor, the researcher should set its vertical and horizontal axes square with the graph paper. The protractor should be moved horizontally until half of the body outline (or half its estimated weight) appears to be on each side of the vertical axis. Then it should be moved vertically until half the body (weight) appears to be above and below the horizontal axis. When it looks right, a pencil is put through the center hole and the protractor rotated slowly back and forth through 90°. If it is at the center of gravity, the masses times the distances of bodily parts bisected by both axes will balance. Note that *the moments of bodily parts as bisected by both axes should remain balanced.* Thus, as the protractor is rotated, the researcher must decide whether a hand way out from the center balances a portion of the posterior close in. The protractor gives some specific basis for making such a judgment. As the protractor is rotated and the portions of the body appear and disappear in diagonal quadrants, if it seems possible that they could balance there is a close approximation of the center of gravity. This can then be marked and the process can be repeated for the remaining body outline. If they do not balance, it is necessary to begin again to find the center of gravity.

Having located the center of gravity at release and contact, the parabola of flight can be constructed. The two points may not be level horizontally, but this makes no difference in determining the successive abscissas which are equidistant per frame regardless of elevation. From the center of gravity at release, a light, horizontal reference line (abscissa) is drawn, and the point at contact is projected perpendicular to this line. The number of frames is counted from release to contact; the horizontal travel is divided by the number of frames, and these are plotted on the abscissa. These are the positions which the center of gravity was above or below in successive frames. The next step is to determine the vertical components per frame (ordinates).

The ordinates depend on the acceleration of gravity, so from here on a table of solutions of $gt^2/2$ for hundredths of second intervals helps materially. If g/2 is 16.1 feet (per second per second), for .01 second it is .0016 *feet;* for .02 second, .0064 *feet;* . . . for .1 second, .161 feet (about 2 inches); for .2 second, .644 feet (about 8 inches), etc. The table is easier to make than it seems, but it must be remembered that digits to the left of the decimal point are feet and those to the right must be converted to inches (by multiplying the decimal fraction by 12). Also, these are actual distances and must be converted according to the scale used in graphing. If a constant subject-to-camera distance was used, a column of scale values for graphing can be made by multiplying the actual values by a conversion factor.

Vertical rise is decelerated and fall is accelerated by gravity alone. Rise from, and fall to, the same level takes equal time. The "high point" of the parabola will fall on the axis at half the time in flight (to the same level), or at half the horizontal distance between release and return to the same level. Unless the difference in elevation at release and contact is appreciable (see below), the high point will fall on the vertical axis half the horizontal distance from release to contact. Half the number of elapsed frames times the interval between frames (in ten-thousandths of a second) will give the half-time of flight (rounded to thousandths, or hundredths, of a second). From this, the distance the body would have fallen during the half-time can be determined from the table and the elevation of the high point plotted on the vertical axis. The high point is the reference point for succeeding calculations.

To locate the position of the center of gravity frame by frame, the half-time is subtracted from the elapsed time of each succeeding frame (working in thousandths of a second and rounding to hundredths). Intervals on the ascending arc will be negative, but the absolute values will give the "distance fallen" from the conversion table. These values are subtracted from the elevation of the high point to find the ordinates for the frames. When this is plotted above the corresponding abscissa (or below if negative), the result is the parabola of flight of the center of gravity frame by frame.

For simplicity in the above description, the center of gravity at release and contact were assumed to be at the same elevation. This is obviously false, partly because elevation is usually lost and rarely gained in flight, but primarily because the camera only by sheer chance catches the subject with his center of gravity at precisely the same elevation at release and contact. Where actual contact occurs more than part of a frame above or below the elevation at release, the frame in which the center of gravity most closely approximates the level at release can be used for the "contact" frame. The duration of flight, and half-time, on this basis will give only a close approximation of the elevation of the high point. The error in using the closer frame for contact is less than half the interval between frames and the half-time halves this error, so the error is often within the possible accuracy of graphing. Regardless of the half-time used, the computed elevation for the center of gravity at release will fall on the reference line, as it should (the computed elevation will be zero), but the center of gravity of the "closer" frame will also fall on the reference line, when it should not. This occurs because the half-time is slightly in error. If the descending arc of the parabola of flight is longer than the ascending, or the duration of flight is considerable, the error between the actual and computed elevation of the center of gravity at contact may be considerable. A closer approximation of the instant of half-time is needed.

The closer approximation is simply a matter of increasing the half-time, if the center of gravity in the observed "closer" frame was above the reference line, or decreasing the half-time if it was below. Or, if the computed elevation at contact is below the observed elevation, the half-time should be increased, and if above, decreased. Briefly, increasing the half-time raises the high point and also the computed elevations (ordinates) on the descending arc; decreasing it lowers them. Depending on the number of frames per second and the interval between frames, increasing or decreasing the half-time may overcompensate and it may be necessary to work in thousandths of a second. With the correct instant of half-time, the parabola of flight will fit the observed contact and release points. With short vertical travel and time in flight, a first approximation will suffice. Longer vertical travel and time in flight require a closer approximation of the instant of half-time *and also* make the errors of rounding thousandths of a second to hundredths appreciable in using the conversion table. Errors of rounding may be minimized by working in thousandths of a second and using linear interpolation between the table values. Finally, although any half-time will give an ordinate of zero at release, so that the computed point coincides with the observed, any change in the half-time requires recomputation of ordinates on the ascending arc as well as the descending arc.

PREPARING ILLUSTRATIONS

The final step in preparing photographic material for a thesis or publication is to make composites. Look back at Figure III. The best sequence was selected and 3 x 5 inch enlargements were made from the 35-millimeter negative. Prints were made of the complete series, even though some pictures made during slow phases of the movement were to be discarded, since the prints are relatively inexpensive. An estimate was made of the *necessary* portion of each print—half to two-thirds being background. From these dimensions, an estimate was made of the number of rows and columns that would fit some multiple of the dimensions of the finished page. Page sizes vary, but some multiple of 7 x 10 inches is fairly standard. From this, the exact height to which each print would be cut was computed so the final rows would be of uniform height. A standard width was also computed, but some prints were cut wider to show specific things, so some had to be cut narrower to keep the length of the row uniform. In this process, a complete sequence of the vital frames should be kept and the sequence filled out with selected frames from the wind-up and follow-through. A neat composite with a maximum of illustrative material requires considerable preliminary planning. The cropped prints can be mounted on poster board with rubber cement, though this should never be used for mounting prints in theses since it only holds for a year or two.

In Figure III, the frame number and elapsed time of each frame were typed on white paper, trimmed, and affixed in the lower left corner. The typewriter used had the large type commonly used in libraries for cataloguing. The composite was 18 x 22 inches in the original. From this an 8 x 10 negative was made. For publication, glossy prints are necessary. But for theses, what is called an "outline special" is sufficient—an 8½ x 11 inch, single weight, nonglossy, contact print. Prices for photographic work vary, but with four or five prints to be reproduced in a thesis, making a composite, labeling it, and rephotographing it is often cheaper than other methods, besides being neat and permanent.

The cheapest and simplest method of preparing graphic material for publication is by "blackline"—a direct positive reproduction process. The material must be prepared in the finished size desired on translucent paper (drafting paper or, for graphs, paper like Dietzgen No. 346). Figure IV was prepared in this way and the labels were typed on with a sheet of carbon *reversed* under the graph paper in order to have sufficient density to print well. If preparation of the graph in the finished size is not convenient, photographic reproduction can be used.

GENERAL SUGGESTIONS

Complete coverage of photography is beyond the limits of this section, but photography has an extensive literature. Several previous discussions of photography as a research tool in physical education have been published (1, 4, 18, 20, 35, 36, 42). However, some additional practical suggestions may help. Photography has no guardian angel for fools and novices, so good results require careful *planning*. A messy *background* is a great distraction. Your theater may have a castoff backdrop that looks badly faded, but will photograph well and provide a neutral background. When photographing the human body in action, the less and the closer fitting the *clothing*, the better. Many photographers prefer indoor conditions and artificial *lighting*. But when the camera must be 30 to 40 feet from the subject to get extensive action without panning (which introduces to-from distortion), sufficient artificial illumination is difficult to manage without blowing fuses. Natural sunlight, outdoors, with the sun at your back, is less artistic, but its use results in sharper negatives because of a smaller aperture and greater depth of focus.

Photographing *swimmers* presents special problems. With the camera just above the water level, so that its axis is perpendicular to the major movement plane of the in-air portion of the stroke, the underwater portion is lost because of total reflection. Raising the camera and shooting down at an angle makes the underwater portion of the stroke visible. But shooting down (or up) underwater with the camera axis not perpendicular to the major movement plane introduces perspective distortion. Also, as the light rays leave the more dense water and enter the less dense air, refraction makes the apparent depth of portions of the body appear much less than their actual depth and waves distort or even obliterate the refracted image. Refraction operates in underwater movies, but a scale object in the major movement plane will make the action measurable and comparable, unless an appreciable to-from component is involved. Unlike their action in running, the arm and leg actions in swimming are rarely exclusively or primarily in the same plane. Consequently, one camera angle will not suffice for the whole stroke, and analysis of parts with later synthesis of the whole is necessary.

For maximum detail and flexibility in processing, *35-millimeter negative* movie film is best. Direct positive and 16-millimeter can be used for observational analysis, but obtaining enlarged prints for analysis and composites may prove difficult. Home movie film (8mm) is not suitable for analytical work. A *telephoto lens* reduces to-from distortion, but requires greater subject-to-camera distance, which may cause problems indoors. Most movie cameras have a circular, rotary *shutter* with an *opening* of about 160 degrees. Manufacturers often have available a shutter with an opening of about 40 degrees. Substitution of the later reduces, in this case by one fourth, the fuzziness of rapidly moving parts and gives sharper pictures. But, since this also reduces the exposure time, compensation must be made in figuring the aperture with a light meter. Finally, it is necessary to remember the major movement plane and to-from distortion, the scale object and timing device, and that the camera must be loaded.

REFERENCES ON PHOTOGRAPHY

1. Adams, Thurston. *Motion Pictures in Physical Education*. New York: Bureau of Publications, Teachers College, Columbia University, 1940. 84 p.
2. Braune, W., and Fischer, Otto. "Uber den Schwerpunkt des menschlichen Körpers." *Sächsiche Akademie der Wissenschaften, Leipzig, Mathematische-Naturwissenschaftliche Klasse. Abhandlungen.* 15:559-672; 1889.
3. Brophy, Kathleen. *A Kinesiological Study of Improvement in Motor Skill.* Doctor's thesis. Madison: University of Wisconsin, 1948.
4. Cureton, T. K. "Elementary Principles and Technique of Cinematographical Analysis as Aids in Athletic Research." *Research Quarterly* 10:3-24; May 1939.
5. Cureton, T. K. "Mechanics of the Broad Jump." *Scholastic Coach* 4:8-9; May 1935.
6. Cureton, T. K. "Mechanics of the High Jump." *Scholastic Coach* 4:9-12; April 1935.

7. CURETON, T. K. "Mechanics of the Shot Put." *Scholastic Coach* 4:7-10; March 1935.
8. CURETON, T. K. "Mechanics of the Track Racing Start." *Scholastic Coach* 4: 14-16; January 1935.
9. CURETON, T. K. "Mechanics of Track Running." *Scholastic Coach* 4:7-10; February 1935.
10. EDGERTON, HAROLD E., and KILLIAN, JAMES R. *Flash.* Boston: Hale, Cushman, and Flint, 1939. 203 p.
11. ELFTMAN, H. "Forces and Energy Changes in the Leg During Walking." *American Journal of Physiology* 125:339-56; 1939.
12. ELFTMAN, H. "The Function of the Muscles in Locomotion." *American Journal of Physiology* 125:357-66; 1939.
13. ELFTMAN, H. "The Work Done by Muscles in Running." *American Journal of Physiology* 129:672-84; 1940.
14. FENN, W. O. "A Cinematographic Study of Sprinters." *Scientific Monthly* 32: 346-54; April 1931.
15. FENN, W. O. "Frictional and Kinetic Factors in the Work of Sprint Running." *American Journal of Physiology* 92:583-611; 1930.
16. FENN, W. O. "Work Against Gravity and Work due to Velocity Changes in Running." *American Journal of Physiology* 93:433-62; 1930.
17. GESELL, A. L. *An Atlas of Infant Behavior.* New Haven: Yale University Press, 1934.
18. GESELL, A. L. "Cinemanalysis: A Method of Behavior Study." *Pedagogical Seminary and Journal of Genetic Psychology* 47:3-16; 1935.
19. GILBRETH, F. B., and GILBRETH, L. M. *Applied Motion Study.* New York: Sturgis and Walton Co., 1917. 220 p.
20. GLASSOW, RUTH B., and BROER, M. "A Convenient Apparatus for the Study of Motion Picture Films." *Research Quarterly* 9:41-48; May 1938.
21. GROVES, W. H. "Mechanical Analysis of Diving." *Research Quarterly* 21: 132-44; May 1950.
22. HUBBARD, A. W., and SENG, C. N. "Visual Movements of Batters." *Research Quarterly* 25:42-57; March 1954.
23. KLOPSTEG, PAUL E., and WILSON, PHILIP D. *Human Limbs and Their Substitutes.* New York: McGraw-Hill Book Co., 1954. 844 p. Chapter 15.
24. KNOLL, W. "Kinematographische Bewegungsstudien an Läufern und Springern." *Leibesübungen und Korperliche Erziehung* 6:4:85-91; 1930.
25. KNOLL, W. "Kinematographische Bewegungsstudien. VIII Mitteillung. Reiter und Pferd." *Arbeitsphysiologie* 8:375-406; 1934-35.
26. KNOLL, W., and BÜSEN, K. "Kinematographische Bewegungsstudien. IV Mitteillung. Wassersprünge." *Arbeitsphysiologie* 6:46-72; 1932.
27. KNOLL, W., and EGGERS, I. "Kinematographische Bewegungsstudien. V Mitteillung. Schwerpunktsbestimmingen under besondere Beruchschtigung der Sprünge." *Arbeitsphysiologie* 7:398-408; 1933. (Explanation of Braune-Fischer Center of gravity method.)
28. KNOLL, W., and FEESER, B. "Kinematographische Bewegungsstudien. VI Mitteillung. Beziehungen zwischen Korpersschwerpunkt und Gleichgewicht." *Arbeitsphysiologie* 8:304-12; 1934-35.
29. KNOLL, W., and MATTHEIS, TH. "Kinematographische Bewegungsstudien: Analyse und Synthese Leichtathletischer Würfe." *Schweizerische Medezinische Wochenschrift* 2:38:925-232; 1931.
30. KNOLL, W., and MORENZ, W. "Der Sportliche Lauf als Vierfusserbewegung." *Arbeitsphysiologie* 5:227-38; 1932.
31. KNOLL, W., and NOCK, K. N. "Kinematographische Bewegungsstudien VII Mitteillung. Der Start zum Kurzstreckenlauf." *Arbeitsphysiologie* 8:335-46; 1934-35.

32. LANOUE, FRED. "Analysis of the Basic Factors in Fancy Diving." *Research Quarterly* 11:102-09; March 1940.
33. *Life* 44:74-75; February 17, 1958.
34. LOCKHART, AILEENE. "A Survey of Devices Used in Measuring Short Time Intervals." *Research Quarterly* 12:757-64; December 1941.
35. LOCKHART, AILEENE. *The Value of the Motion Picture as an Instructional Device in Learning a Motor Skill.* Doctor's thesis. Madison: University of Wisconsin, 1942. 98 p.
36. MACEWAN, CHARLOTTE, and HOWE, EUGENE. "An Objective Method of Grading Posture." *Research Quarterly* 3:144-57; October 1932.
37. MCGRAW, M. B. *A Study of Jimmy and Johnny.* New York: D. Appleton and Co., 1935.
38. MAREY, E. J. *Animal Mechanism:* A Treatise on Terrestrial and Aerial Locomotion. New York: D. Appleton and Co., 1874. 283 p.
39. MAREY, E. J. "The History of Chronophotography." *Smithsonian Institution Annual Reports* 56:317-40; 1902.
40. MAREY, E. J. *Movement (Le Mouvement).* Translated by Eric Pritchard. New York: D. Appleton and Co., 1895. 323 p.
41. MUYBRIDGE, EADWEARD. *The Human Figure in Motion.* London: Chapman and Hall, 1901. 280 p. (Republished in 1955 with an introduction by Robert Taft.)
42. PRIEBE, ROY E., and BURTON, WILLIAM B. "The Slow Motion Picture as a Coaching Device." *School Review* 47:192-98; 1939.
43. SCHMITH, O. "Zur Physiologie von Sportbewegung. I Mitteillung. Fragestellung. Untersuchungsmethodik." *Arbeitsphysiologie* 5:370-83; 1932.
44. SCHMITH, O. "Zur Physiologie von Sportbewegung. II Mitteillung. Der Hürdenlauf." *Arbeitsphysiologie* 5:394-423; 1932.
45. SCHMITH, O. "Zur Physiologie von Sportbewegung. III Mitteillung. Der Leichtathletische Lauf." *Arbeitsphysiologie* 7:83-107; 1933.
46. SHERMAN, HOYT L. *Drawing by Seeing.* New York: Hinds, Hayden, and Eldridge, 1947.
47. SULLIVAN, R. N. *The Forward Somersault on the Parallel Bars.* Master's thesis. Urbana: University of Illinois, 1953. 68 p.

CHAPTER 6

Laboratory Planning and Instrumentation

FRANK SILLS

THERE ARE A NUMBER OF ESTABLISHED LABORATORIES IN PHYSICAL education that have been very productive. The majority of these laboratories have been developed through the efforts of individuals who have had the training and the necessary interest to conduct laboratory research. It is now apparent that many more schools and universities are interested in sponsoring research laboratories in their physical education departments.

Fundamentally, a laboratory in physical education is not very different from that in any other area and, for this reason, it is wise to consider accepted principles underlying the development of a research laboratory. It is hoped that the principles cited at the end of this chapter will guide persons intending to set up laboratories and that these principles may be used on the basis of their relative merits in respect to a particular laboratory.

PLANNING THE LABORATORY

Many physical educators assume that the cost of a research laboratory is prohibitive and do not, therefore, consider developing one. A laboratory can be developed, and probably should be, in those institutions that have a graduate curriculum in physical education. In most cases, the initiative must come from faculty members who desire to work in a laboratory situation. The simplest way to begin, in the author's opinion, is to develop a number of research projects which require inexpensive laboratory equipment. Once the laboratory is underway, it is probable that the school administrator will encourage its development

Once a few items of equipment are purchased or are constructed by the members of the staff, the laboratory is underway. It is now a problem of setting up additional research projects utilizing the original equipment and obtaining additional equipment as needed. By small additions over a period of a few years, a well-equipped laboratory will evolve. Just how elaborate the laboratory will be is a matter for the administrator and staff to decide. The production of good research in selected areas may attract the attention of research foundations or other sponsoring organizations, or may serve as a basis for requesting financial aid from the school's budget.

Those who are fortunate enough to obtain a large grant to start a laboratory should not buy equipment and supplies indiscriminately. To purchase an electromyograph for a laboratory simply because some other school has been doing research with this instrument is foolish because there may never be anyone in the laboratory interested in this type of research. The initial grant should be used only as needed, and future needs should be anticipated very carefully. It is wise to remember that the laboratory which produces a considerable amount of valuable research with a limited amount of eqiupment is a better laboratory than the one which is elaborately equipped but used sparingly. It is very important, then, that equipment and instruments be purchased or built as the demand for them arises.

One of the more serious mistakes commonly made in a laboratory is the purchase of expensive equipment without an adequate maintenance budget. If the repair and maintenance of equipment can be taken care of by the personnel who regularly work in the laboratory, the item in the budget for this purpose is small than is the case when "outside" technicians must be employed. Some schools are fortunate enough to be able to employ an engineering student for part-time work, while others are fortunate enough to have a medical school nearby where serological and other tests may be obtained at a minimum cost.

All of these factors, and many more, must be taken into consideration in the planning of a laboratory.

INSTRUMENTATION

The equipment and instruments described here are those which have been used in the field of physical education during the past several decades. It is impractical to include a detailed description of every item, and the author has attempted to select only those that are representative of different types of research.[1]

The presentation of the instruments is divided into ten categories, each category representing one or more phases of human performance: speed, strength and force, reaction time, neuromuscular tremor, kinesthesis, balance, muscle voltage, metabolism, circulo-respiratory endurance, and selected factors of physical performance.

Since many of the instruments used for research in physical education have not been given specific titles, the presentations for each category are made according to the applications of the instruments. For this reason, a heading such as "Speed of Rotary Arm Movement" may appear twice, and two different instruments for measuring this particular physical performance may be presented. Each presentation is divided into Basic Components, Application, and Source.

Measurement of Speed

1. Speed of Rotary Arm Movement—A

Basic Components: gear reduction box, cam microswitch, and electric timer.

Application: This apparatus may be used to measure the speed of rotary movements of the arm. By turning the handle, the subject causes a cam, which is set at a 1-24 ratio, to rotate. The cam completes the electric circuit which starts the clock. After 24 complete revolutions of the arm, the cam will break the electrical circuit, stopping the clock.

Source: Zorbas, William S., and Karpovich, Peter V. "The Effect of Weight Lifting Upon the Speed of Muscular Contraction." *Research Quarterly* 22:145-48; May 1951.

2. Speed of Rotary Arm Movement—B

Basic Components: bicycle crank with a radius of 7¼ inches mounted in a frame and attached to a strong upright on the wall, the axis of the crank 58 inches from the floor; an electric counter which can be read at 15-second intervals.

[1] A list of the companies from which equipment of this type may be purchased is available upon request to the Research Council of the American Association for Health, Physical Education, and Recreation.

Application: This apparatus is used to test speed of movement involving the use of the arms and shoulders. The subject takes hold of the crank with both hands and turns the crank at maximum speed. The number of turns is recorded on the electric counter and read at 15-second intervals. If desired, a fatigue curve may be obtained by plotting the rate of movement in successive 15-second time periods.

Source: Wilkin, Bruce M. "The Effect of Weight Training on Speed of Movement." *Research Quarterly* 23:361-69; October 1952.

3. Speed of Arm Movement—A

Basic Components: two movable posts covered with sponge rubber, and chronoscope.

Application: This apparatus may be used to measure the speed of movement of the arm or other parts of the body from one point to another. One application is to place the posts a given distance apart, so that releasing the hand from one post will close the circuit to the standard timer and striking the second post will open the circuit to the standard timer. The time is recorded on the chronoscope.

Source: Rasch, Philip J. "Relationship of Arm Strength, Weight, and Length to Speed of Arm Movement." *Research Quarterly* 25:328-32; October 1954.

4. Speed of Arm Movement—B

Basic Components: synchronized chronoscope graduated in .01-second units, make and break switches, and target.

Application: This apparatus may be used to measure the forward velocity of the dominant hand over a distance of 11 inches. The apparatus may be modified to meet the demands of various kinds of experiments.

The subject places his hand on a switch which, when released, starts the chronoscope. He then moves his hand forward a distance of 11 inches, striking the target, which stops the chronoscope.

Source: Pierson, Wm. R. "Comparison of Fencers and Non-Fencers by Psychomotor, Space Perception, and Anthropometric Measures." *Research Quarterly* 27: 90-96; March 1956.

5. Speed of Body Movement

Basic Components: control box with relay, chronoscope make-switch box, and break-switch box.

Application: This apparatus may be used to measure speed of body movements. The make switch and break switch are set in separate units which can be moved to any desired location. The apparatus may be used to measure speed of leg movement, arm movement, trunk movement, or movement of the entire body from one location to another. The subject releases a snap switch which makes the contact starting the chronoscope, and then strikes the second switch which breaks the contact with the chronoscope.

Source: Sills, Frank D. Physical Education Research Laboratory, State University of Iowa.

6. Speed of Co-Ordinated Movement

Basic Components: an operating key comprised of a repositioned silent operating switch, a thyratron electronic tube, light and sound stimuli, an inverted amplifier output transformer with a low voltage supply, skin electrodes, and a chronograph.

Application: This apparatus may be used to measure single reaction time and speed of co-ordinate movements. An adjustable electronic delay circuit provides a means for administering a mild electric shock when the subject makes a slow response. Either light or sound stimuli may be used.

Source: Henry, Franklin M. "Increase in the Speed of Movement by Motivation and Transfer of Motivated Improvement." *Research Quarterly* 22:219-28; May 1951.

7. Speed of Running

Basic Components: a chronograph consisting of a constant speed phonograph motor with a 50-tooth gear turned by a telechron motor, a recording pen, starting box, and contact gates which make the circuit, causing a recording to be made on the chronograph.

Application: This apparatus may be used to measure the velocity of a runner at various distances up to 50 yards. The switches on the 10 gates placed at 5-yard intervals are hooked up in series, as is a clap board with electrical contacts that provide the starting signal. The runner's time for each 5 yards up to 50 yards is recorded on the chronograph as he runs through each successive gate.

Source: Henry, Franklin M., and Trafton, Irving R. "The Velocity Curve of Sprint Running." *Research Quarterly* 22:409-22; December 1951.

8. Speed of Repetitive Movement

Basic Components: specially constructed platform with one half rigged in such a way that a spring lifts platform when foot is removed, electrical contacts on platform, and Veeder-Root Electric Counter.

Application: This apparatus may be used to measure speed of running in place. The subject stands on the platform. Each time he raises his left foot he breaks contact, and each time he lowers his foot he makes a contact activating the electric counter. The number of steps which the subject takes in a given period of time may be accurately determined. Two counters may be used with both the right and left halves of the platform hinged. An electric timer may be included in the circuit, so that throwing the starting switch will turn on both the timer and counter at the same time. At the end of ten seconds, both the contact to the counter and electric timer may be broken.

This apparatus may also be used in a vertical position to measure speed of punching.

Source: Sills, Frank D., and O'Riley, Vernon E. "Comparative Effects of Rest, Exercise, and Cold Spray Upon Performance in Spot Running." *Research Quarterly* 27:217-19; May 1956.

9. Velocity Measurement of Baseball

Basic Components: baseball partially coated with conducting silver electrodes to be attached to fingers, electroswitches, standard electric timer, and a speaker unit.

Application: This device is used to measure the velocity of a pitched ball. When the pitcher holds the ball with the electrodes across the coated area, the electrical circuit is opened. When the ball is released, the electric timer begins running and, when the ball is caught, the sound waves created are picked up by a speaker unit so that the impulse from the sound wave is used to stop the timer.

Source: Slater-Hammel, A. T., and Andres, E. H. "Velocity Measurement of Fast Balls and Curve Balls." *Research Quarterly* 23:95-97; March 1952.

10. Velocity of Bodily Movements and Projectiles

Basic Components: cathode-ray oscillograph, wide range oscillator, transformer, condenser, resistors, 35mm camera, camera tripod, and contact points for circuit as desired.

Application: This apparatus may be used to measure a single response or multiple responses. It is possible to impress a time wave on the cathode-ray oscillograph and to initiate the movement of the ray across the face of the CRO. As the ray moves across the face of the CRO, it is possible to interrupt it by means of completing electrical circuits in parallel.

The apparatus may be used effectively to measure velocity, for problems similar to that presented under 4. Speed of Arm Movement—B above, and also to measure the velocity of objects traveling at extremely high velocities such as thrown balls, batted balls, and similar projectiles. For measurements of the latter type, it is necessary to use either a sound pick-up or some type of photo-electric screen. In the measurement of velocity of running, it is possible to record measurements in fiftieths or hundredths of a second; and for measurements of the velocity of thrown or batted balls, it is possible to measure to an accuracy of 5/1000 of a second.

Source: Sills, Frank D. Physical Education Research Laboratory, State University of Iowa.

Measurement of Strength and Force

1. Resistance and Propulsion in Swimming

Basic Components: a one-horsepower, two-phase, 6-cycle, 220-volt motor, rated at 1750 revolutions per minute; V-belt, steel shaft, 5-step pulley, and heavy cord for towing swimmer; kymograph and spring scale to measure force; and a pacing device consisting of a synchronous motor, storage battery, pulley system, and automobile horn. (A tape recorder with speaker may also be used as a pacing device.)

Application: This apparatus may be used to investigate the problem of water resistance and propulsion in swimming. The subject wears a belt with a small steel ring to which the heavy cord may be fastened. This cord extends to either end of a steel shaft. The shaft is controlled by the motor, and the revolutions may be varied according to which one of the five pulleys is employed. The force exerted by the swimmer upon the apparatus which is suspended is measured by means of a kymograph. An electrically heated stylus extends from the arm of the scale, making a continuous mark on the waxed paper of the kymograph.

Source: Alley, Louis E. "An Analysis of Water Resistance and Propulsion in Swimming the Crawl Stroke." *Research Quarterly* 23:253-70; October 1952.

2. Forces Exerted in Swimming

Basic Components: the same components as given under 1. Resistance and Propulsion in Swimming above. Instead of using a suspended apparatus and spring scale, two steel beams which permit less than 1/100-inch movement under loads of 50 pounds are attached to the sides of the swimming pool. Four 500-ohm bobbin-type strain gauges, a Hathaway MRC-12 strain gauge control unit (which serves as an amplifier), and a Hathaway S-14-C with recording paper and a recording galvanometer.

Application: This equipment may be used to measure drag and effective propulsive force when the swimmer is pulled through the water or when he swims against resistance. The strain gauges measure the amount of deflection imparted to the beams. The deflection is amplified and recorded so that the amount of force exerted by the swimmer is known. A timing wave of 60 cycles per second is marked on the recording paper, and a pacing device, like the one described under 1. Resistance and Propulsion in Swimming above, is used.

Source: Counsilman, James E. "Forces in Swimming Two Types of Crawl Stroke." *Research Quarterly* 26:127-39; May 1955.

3. Charging Force

Basic Components: microphone, amplifier, clock, I-beam on which a padded dummy is mounted, tension springs, calibration spring, microswitch, roller bracket, and recoil spring.

Application: This instrument may be used to measure the speed and force of a football charge. When the snap signal is given the impulse is picked up by the microphone, amplified, and used to start the electric clock. When the subject strikes the dummy, which is attached to the I-beam, the movement of the I-beam is determined by means of a pointer which is attached to it. This pointer indicates the amount of force exerted and will remain in place until it is reset manually.

Source: Elbel, Edwin R., and others. "Measuring Speed and Force of Charge of Football Players." *Research Quarterly* 23:295-300; October 1952.

4. Force and Time Factors Involved in Sprint Start

Basic Components: the same as indicated in 4. Speed of Arm Movement—B under the section on Measurement of Speed; in addition, starting blocks consisting of wooden face plates mounted on a metal carriage, roller bearings, two rack and pinion combinations, and a kymograph.

Application: In addition to the recording of velocity, the amounts of force applied on the rear foot plate and front foot plate are measured, in pounds, by means of the sliding carriage as it presses against the coil spring. The time characteristics are recorded by means of an electromagnet with the stations at 5-yard intervals as indicated under 4. Speed of Arm Movement—B in the section on Measurement of Speed.

Source: Henry, Franklin M. "Force-Time Characteristics of the Sprint Start." *Research Quarterly* 23:301-; October 1952.

5. Muscle Strength

Basic Components: cable tensiometer, cable, and attachments.

Application: The cable tensiometer, which is manufactured by the Pacific Scientific Company of Pasadena, California, may be used to measure the strengths of numerous muscle groups. It is left to the experimenter to determine the manner in which the 1/16-inch cable should be attached to the subject and to a fixed point. The tension on the cable is measured by running it through the tensiometer. One of the advantages of using the tensiometer is that the excursion through which the subject may move a particular body part is limited, thereby eliminating a possible source of error.

Source: Clarke, H. Harrison. "Comparison of Instruments for Recording Muscle Strength." *Research Quarterly* 25:398-411; December 1954.

6. Muscle Strength and Muscle Endurance—A

Basic Components: grip dynamometer comprised of aluminum grips, strain gauge (containing a Wheatstone bridge so that an increase in resistance on two wires of the bridge and a slight decrease on the second two cause imbalance), an AC amplifier, and a 5-milliampere DC Esterline-Angus Recorder.

Application: This instrument may be used to measure both isometric and isotonic work. The subject's maximum application of force during a specified period of time may be recorded to determine muscle endurance. This type of instrument may be modified so that it may be used to measure the strength and strength endurances of numerous muscle groups.

Source: Thompson, Clem W. "Some Physiologic Effects of Isometric and Isotonic Work in Man." *Research Quarterly* 25:476-82; December 1954.

7. Muscle Strength and Muscle Endurance—B

Basic Components: Baldwin SR-4 Type U-1 load cell, amplifier, and recorder.

Application: This apparatus is an adaptation of that described under 6 just above. It is used to measure back and leg strengths and may be calibrated from 600 to 3000 pounds. The apparatus may also be modified to measure abdominal strength and strengths of muscles not so strong as the back and leg muscles.

Source: Tuttle, W. W., and others. "Relation of Maximum Back and Leg Strength Endurance." *Research Quarterly* 26:96-106; March 1955.

8. Muscle Strength and Muscle Endurance—C

Basic Components: the components described by Kelso and Hellebrandt, which include the following: a wheel and axle, two counters, a distance meter, kymograph, mechanical chronograph driven by 6rpm voltage reduction and insulating transformer, and apparatus for connecting subject with ergograph.

Application: The ergograph is used to study repetitious movements rather than a single movement. It provides a method for measuring fatigue in terms of the number of times a given resistance can be moved. The most recent modification of the ergograph is that by Hellebrandt and Kelso in which a wheel and axle have replaced the sling previously used by Mosso. The wheel and axle made it possible to keep the angle of pull the same throughout the entire movement.

The apparatus can also be used in doing repetitious bouts of a set number of contractions against progressively increasing resistances. This may be done by giving a subject a short rest between bouts, and adding resistances as desired. The Hellebrandt-Kelso Ergograph includes an electromagnetic signal for indicating the number of contractions performed, and provides control of the load by means of two vertical rods with a minimum of friction. It eliminates overhanging parts found in previous models of ergographs, and also has a distance meter for recording the accumulative height that the resistance is lifted.

Source: Clarke, H. Harrison. "Recent Advances in Measurement and Understanding of Volitional Muscular Strength." *Research Quarterly* 27:263-75; October 1956.

9. Abdominal Muscle Strength and Endurance

Basic Components: special table (Ann Arbor Instrument Works, 725 Packard, Ann Arbor, Michigan) of steel construction with padded plastic surface, 6½ feet long by 3 feet wide; resistance coil with a movable lever, resistance rod, and crank; bar attached to lever that crosses subject's chest; kymograph; timing device; and resistance indicator.

Application: Apparatus may be used in the testing of abdominal muscle strength and endurance. The subject lies on the table in a hook lying position with a rigid bar tight against the chest, and two inches below shoulder level. The resistance indicator is set to register 20 pounds, and the subject raises his trunk off the table as far as possible. The trunk is held in this position for 30 seconds. The highest point on the curve drawn on the kymograph is used in computing the strength of the abdominal muscles, while the area under the line drawn on the kymograph is measured by a planimeter to determine the strength endurance of abdominal muscles.

Source: Walters, C. Etta, and Harris, Ruth W. "An Apparatus for Measuring Abdominal Strength and Endurance." *Physical Therapy Review* Vol. 33; September 1953.

Measurement of Reaction Time

1. Football Charging Time

Basic Components: chronoscope, target attached to the extension of a hinged typed circuit breaker, buzzer, leather helmet, and mattress.

Application: This apparatus may be used to determine the response time of a charge similar to that used by a football player. The subject takes a three-point stance and, in response to the buzzer signal, charges the target which he strikes with his head. The distance between the target and the subject's head is 12 inches. Following his charge, the subject may fall forward on the mattress, which helps break the force of his fall.

Source: Manolis, Gus G. "Relation of Charging Time to Blocking Performance in Football." *Research Quarterly* 26:170-78; May 1955.

2. Total Body Reaction Time

Basic Components: two metal contact plates, light stimulus, chronoscope, platform, microswitch, and visual stimulus consisting of a 105-125 split-plate, neon glow lamp.

Application: The total body reaction time to be measured by this apparatus is done in the following way: When the visual stimulus is given, the subject steps forward. When the left light comes on, the subject steps off with the left foot and when the right light comes on, the subject steps forward with the right foot.

Source: Slater-Hammel, A. T. "Initial Body Position and Total Body Reaction Time." *Research Quarterly* 24:191-96; March 1953.

3. Reaction Time from Visual Stimulus

Basic Components: an angle-iron framework, wooden base, metal plate, a marble or a rubber ball.

Application: This apparatus may be used to measure simple reaction time as determined by movement of the hand. The subject places his hands at varying distances from the metal plate behind which is hidden a small marble or ball. When the object drops into the subject's field of vision, he moves his hand as quickly as possible from the hand rest. The distance from the hand rest to the marble (behind metal plate) will determine the time lapse. By varying the distance of the hand from the hidden object, the reaction time of the subject is determined.

Source: Slater-Hammel, A. T. "An Inexpensive Gravity Reaction Time Device." *Research Quarterly* 25:218-21; May 1954.

4. Reaction Time from Peripheral Stimuli

Basic Components: neon glow lamps mounted around an arc with a radius of 58 centimeters, headrest mounted at center of arc described by lights and adjustable in 3 dimensions, response key for subject, control switch, chronoscope, and 300-cycle-per-second tone to mask external disturbances.

Application: This apparatus is useful in determining reaction time to light stimuli located at various positions in the peripheral field. The subject responds to whatever light is activated by the experimenter.

Source: Slater-Hammel, A. T. "Reaction Time to Light Stimuli in the Peripheral Visual Field." *Research Quarterly* 26:82-87; March 1955.

5. Reaction Time from Movement Stimulus

Basic Components: The components are similar to those described by V. H. Denenberg in the *American Journal of Psychology* in 1953. They consist of an electro-magnet energized by direct current, experimenter's switch, soft iron bar riveted to a leather wrist band, telegraph key, and chronoscope.

Application: This instrument is used to measure reaction time as initiated by movement of a part of the body. In using this apparatus, the subject abducts his left arm, to which is attached the iron bar that comes in contact with the electro-magnet. The magnet holds the subject's arm in abduction until the experimenter's switch makes the magnet inoperative. At the same time, the chronoscope is started and will continue to run until the subject removes the fingers of his opposite hand from a telegraph key.

Source: Slater-Hammel, A. T. "Comparisons of Reaction Time Measures to Stimulus and Arm Movement." *Research Quarterly* 26:470-79; December 1955.

6. Choice Response Time

Basic Components: control box consisting of relays, light stimuli, and control switches for operator; target switches (adjustable as desired by experimenter); and chronoscope.

Application: This apparatus may be used to determine the response time of an individual when a choice has to be made. There are four stimulus lights which indicate the targets which the subject must touch. The targets may be placed at any position desired. When a light stimulus appears, the chronoscope circuit is closed; then, when the subject strikes the correct target, the chronoscope circuit is opened.

Source: Wendler, A. J. Physical Education Research Laboratory, State University of Iowa.

7. Reaction Time and Multiple Response Time

Basic Components: same as for preceding test plus a second chronoscope in a separate circuit to measure reaction time.

Application: When the light stimulus appears, the first chronoscope circuit is closed. Then, when the subject removes his hand or foot from the reaction-time key, the first chronoscope circuit is opened so that reaction time is measured. Simultaneously with the removal of the hand or foot from the reaction time switch, the second chronoscope circuit is closed. This second circuit will be opened when the subject strikes the appropriate target so that response time may be measured.

Source: Sills, Frank D. Physical Education Research Laboratory, State University of Iowa.

8. Reaction Time from Auditory Stimulus

Basic Components: impulse counter and interval timer, telegraph key, and a hand switch.

Application: This apparatus may be used for measuring the combination of reaction time and response time. The subject responds to a stimulus bell, which also starts the timer. When the subject strikes the target, it stops the ringing of the bell and the timer. The same type of hook-up may be used to measure movement of the entire body in response to a stimulus.

Source: Atwell, Wm. O., and Elbel, Edwin R. "Reaction Time of Male High School Students and Fourteen-Seventeen-Year Age Groups." *Research Quarterly* 19:22-29; March 1948.

Measurement of Neuromuscular Tremor

1. Neuromuscular Tremor Amplitude

Basic Components: a strain gauge, amplifier, and recorder.

Application: This apparatus may be used to measure the magnitude of neuromuscular tremor. Neuromuscular tremor has been shown to be a sensitive measure of the effects of stress upon physiologic responses. The subject's static tremor is recorded from the index finger of an outstretched arm. The finger barely touches the activated pin of the strain gauge which is supported on a ring stand. The pressure on the activated pin changes the potential in a Wheatstone bridge, and the change in potential is recorded. The voltage is amplified prior to the time that it is recorded on graph paper.

Source: Mitchem, John C., and Tuttle, W. W. "Influence of Exercises, Emotional Stress, and Age on Static Muscular Tremor Magnitude." *Research Quarterly* 25: 65-74; March 1954.

2. Neuromuscular Tremor Duration

Basic Components: a chronoscope, stylus, metal plate with a target hole.

Application: This apparatus may be used to measure neuromuscular tremor in a manner similar to that in the preceding discussion. However, instead of using a strain gauge, amplifier, and kymograph, an electric timer records the number of seconds during which the subject is unable to keep the metal stylus from contacting the edge of the target hole. The opening used in the metal plate is 0.147 inches, and the tip of the stylus has a diameter of 0.055 inches.

Source: Slater-Hammel, A. T. "Influence in Order of Exercise Bouts Upon Neuromuscular Tremor." *Research Quarterly* 26:88-95; March 1955.

Measurement of Kinesthesis

1. Kinesthetic Perception and Adjustment

Basic Components: two vertical shafts, mounted on a stand, which pivot freely; tension spring connecting the two vertical shafts; cam; motor; speed-reducing gear box; a chronograph specially designed from a monodrum; and a variable range, high speed synchronized power unit.

Application: This apparatus is used to measure constant pressure and constant position: (a) To measure constant pressure, the subject must push with the same pressure against a pad on a vertical shaft while the pressure changes from the influence of the cam; and (b) For the constant position test, the subject increases or decreases his pressure to meet the changing pressure caused by the cam.

Source: Henry, Franklin M. "Dynamic Kinesthetic Perception and Adjustment." *Research Quarterly* 24:176-87; May 1953.

2. Kinesthetic Perception

Basic Components: nonshatterable rubber goggles with adjustable lenses blacked out with black poster paint, four finger pointers cut from thin copper sheet, reflector with several No. 2 photofloods, six-foot-high support for flood lamp, five-foot-square measuring board constructed from quarter-inch plywood with a 28-inch radius circle on the face of the board, and radii constructed at each degree, system of hooks and pulleys to adjust board to varying heights of subjects, and guidelines on floor for replacement of subject's feet.

Application: This apparatus is used to measure kinesthetic perception. It may be used to determine the ability of an individual to place his arm in a given position. The individual's arm is placed in position by examiner. The subject is then instructed to duplicate the position. The subject's "score" is then read from the measuring board. Facial tissue is inserted between the subject's eyes and dark glasses to ensure an adequate blindfold.

Source: Phillips, Marjorie, and Summers, Dean. "Relation of Kinesthetic to Motor Learning." *Research Quarterly* 25:456-69; December 1954.

Measurement of Balance

1. Body Sway

Basic Components: spring-driven and governor-controlled kymograph, a pulley system, and an adjustable cap.

Application: This apparatus may be used to measure body sway. The subject's body sway is transmitted from the helmet, which he wears on his head, by means of a pulley system to the ink writer kymograph, which in turn records the movements of the subject into equivalent vertical lines on adding machine paper.

Source: White, Delmer V. "Static Ataxia in Relation to Physical Fitness." *Research Quarterly* 22:92-101; March 1951.

2. Balance—A

Basic Components: stabilometer platform and chronograph.

Application: The subject takes a kneeling position similar to that of referee's position in wrestling. He attempts to maintain balance as nearly as possible and a line is drawn on the chronograph by means of a pen that is connected to the stabilometer platform.

Source: Mumby, Hugh H. "Kinesthetic Acuity and Balance Related to Wrestling Ability." *Research Quarterly* 24:327-34; October 1953.

3. Balance—B

Basic Components: a 40-inch by 7½-inch by 1½-inch teeter board with metal axle, pneumatic stops, moving contact arm, a bank of five white lamps, and a bank of five stimulus lamps; a stepper relay operated by an electric time delay switch; and two standard electric clocks.

Application: This apparatus is used to measure an individual's balance by means of a stimulus response procedure. The subject will see a red light on the stimulus display, and it is his task to move the teeter board in such a position that the white light immediately below the red light comes on. The total test consists of having the subject change his position continuously, bringing on the matching white lights as the red stimulus lights are lighted by the movement of the stepper relay. This apparatus is similar to the one used by Reynolds, and the circuits are designed to perform exactly as those described by Reynolds. The apparatus is useful in measuring balance in response to sequential presentation of stimuli.

Sources: Reynolds, Bradley. "Correlation Between Two Psychomotor Tests as a Function of Practice on the First." *Journal of Experimental Psychology* 43:341-48; 1952.

Slater-Hammel, A. T. "Performance of Selected Groups of Male College Students on the Reynolds Balance Test." *Research Quarterly* 27:347-52; October 1956.

4. Balance—C

Basic Components: wooden platform 22 inches long and 20 inches wide, mounted on a steel shaft which is anchored on ball-bearing pivots; microswitches; control switch; and chronoscope.

Application: This balance platform may be used to determine dynamic balance over any given period of time. A stop watch may be used to measure the total time involved in the test, while the chronoscope will indicate how much of the time the individual is on balance. The excursion of the platform can be adjusted so that the distance required for maintaining balance can be varied. Each time the subject loses his balance to either the right or the left, the microswitch under the platform makes contact, closing the circuit to the chronoscope. This apparatus has been modified and used to determine balance in a sitting position.

Source: Wendler, A. J. Physical Education Research Laboratory, State University of Iowa.

Measurement of Muscle Voltage

1. Electromyograph

Basic Components: The components described here are found in the Grass Model III-D electroencephalograph—electrode board, calibration controls, electrode selector switch, pre-amplifiers, 6-volt A battery, 45-volt B battery, and skin electrodes.

Application: This instrument may be used for electromyography, electroencephalography, and electrocardiography. Application as an electromyograph will be considered here. It may be used in kinesiological analyses and analyses of pathological conditions. Skin electrodes are attached to the surface immediately over the muscle to be evaluated or, if desired, needle electrodes may be inserted into the muscle. The electrodes which are placed over the muscle to be evaluated will "pick up" the microvoltage generated in the muscle tissue when the muscle contracts. This voltage is amplified and recorded by the ink-writer assembly. With this particular apparatus it is possible to use from one to eight channels, so that eight different muscle areas may be selected for evaluation simultaneously. Interpretation of the recording is based upon the vertical deflections of the ink-writing pens and will be dependent upon the calibration of the apparatus prior to the time the recording is made.

Electromyography is useful in determining the relative contribution of a muscle to a particular movement, and may be used to determine the amount of muscle activity that is present in injury or paralysis. The most common applications that have been made in physical education are those related to basic anatomical movements and to the performance of sports skills.

Another type of electromyograph is that which has a cathode-ray oscillograph, such as the Meditron, for interpreting the microvoltage of a muscle. This instrument operates in a manner similar to the one that uses the ink-writing unit. However, it produces a wave form that is more representative of the muscle voltage developed. The difficulty in using this apparatus is to procure permanent records. The investigator must use a motion picture camera especially adapted to the cathode-ray oscillograph if he desires a permanent record.

It is possible to take a single exposure of the wave form on the scope with a 35mm camera. This provides a permanent record but not the same kind of continuous record that may be procured by means of an ink-writing assembly. One of the advantages of this type of machine is that the investigator may observe the oscillograph while the subject is undergoing a test and may obtain a direct reading of the microvoltage the subject is developing in a muscle. The Meditron electromyograph has a tape recorder that will record the muscle's activity, and which may be played back so that the sound of the muscle "firing" may be heard. At the same time, the wave form developed by the muscle may be observed on the cathode-ray oscillograph when the recording is played back. It would seem that the ink-writing unit has some disadvantages, while the cathode-ray oscillograph has others.

Sources: Sigerseth, P. O., and McCloy, C. H. "Electromyographic Study of Selected Muscles Involved in Movements of Upper Arm at Scapulohumeral Joint." *Research Quarterly* 27:409-17; December 1956.
Sills, Frank D., and Olson, Arne L. "Action Potentials in Unexercised Arm When Opposite Arm Is Exercised." *Research Quarterly* 29:213-21; May 1958.

2. Electrocardiograph

Basic Components: skin electrodes, amplification system, and recording system.

Application: Surface electrodes are placed in certain conventional locations reached by general agreement. The electrocardiogram shows during each cardiograph cycle three positive deflections or waves and two negative ones. The first positive wave corresponds to the spread of excitation in the auricles and is known as the P Wave. Subsequent deflections follow the letter P alphabetically and are known as QRST Waves. The R and T waves are positive and the Q and S are negative. Three leads commonly used are the *right arm* to *left arm*, *right arm* to *left leg*, and *left arm* to left leg. These leads give an over-all view of the electric potentials that are developed by the heart beat. The variation in the electrocardiographic recording tells whether the impulse originated and spread along normal or abnormal paths at normal or abnormal speeds.

Source: Houssay, Bernardo A., and others. *Human Physiology.* New York: McGraw-Hill Book Co., 1951, p. 114-25.

Measurement of Metabolism

1. Gas Analysis

Basic Components: Douglas bag, gas meter, two-way breathing valve, carbon dioxide and oxygen gas analysis apparatus.

Application: This apparatus is for measurement of metabolism during rest and work. Expired air is collected in Douglas bag, and volume is measured by means of gas meter. Sample of air is analyzed to obtain the respiratory quotient. From this information, the energy cost of work may be computed.

Source: Bailey, Cameron B. "Apparatus Used in the Estimation of Basal Metabolism." *Journal of Laboratory and Clinical Medicine* 6:657-79; September 1928.

2. Respirometer

Basic Components: a counterbalanced spirometer bell, ventilometer, kymograph, and appropriate accessories, including writing pens, valves, hoses, and switches.

Application: This apparatus may be used to measure tidal volume, vital capacity, inspiratory and expiratory reserve volume, inspiratory capacity, functional residual capacity, and total lung capacity. The Collins respirometer is an example of this type of instrument and is commonly used to measure oxygen consumption either at rest or during work. From this information, the energy cost of doing work may be determined. An estimated respiratory quotient must be used since the expired air is not analyzed.

Source: Best, Charles H., and Taylor, Norma B. *The Physiological Basis of Medical Practice.* Baltimore: The Williams & Wilkins Co., 1945.

Measurement of Circulo-respiratory Endurance

1. Bicycle Ergometer

Basic Components: electrodynamic brake, amplifier, Esterline-Angus recorder, and stationary bicycle.

Application: This apparatus is used to determine the amount of work which an individual can do in a given period of time and is directly associated with measurements of muscle endurance and circulo-respiratory endurance. It is possible to vary the current to the electrodynamic brake in order to increase or decrease the resistance. When the amperage to the brake is increased, the subject has to perform a greater amount of work; as the amperage is decreased, the subject does a lesser amount of work. The work which the subject performs in pedaling the bike is amplified and recorded on an Esterline-Angus recorder. It is possible to measure all-out work or to measure the ability to perform with a given work load over a pre-determined period of time.

A less costly friction-type bicycle may be used if desired.

Sources: Karpovich, Peter V. "A Frictional Bicycle Ergometer." *Research Quarterly,* 21:210-15; October 1950.

Tuttle, W. W., and Wendler, A. J. "The Construction, Calibration and Use of an Alternating Current Electrodynamic Brake Bicycle Ergometer." *Journal of Laboratory and Clinical Medicine* 30:173-83; February 1945.

2. Treadmill

Basic Components: motor-driven treadmill adjustable to various grades and velocities.

Application: This apparatus is useful in the same types of research as the bicycle ergometer. The work which the subject performs may be increased by accelerating the treadmill or by increasing the incline. Characteristic recordings taken from a subject during work on the treadmill include heart rate, respiratory volume, blood pressure, oxygen consumption, and temperature variations.

Measurement of Selected Factors Associated with Physical Performance

1. Co-ordination

Basic Components: a 3½-foot triangle target and a foot-square target; copper discs, 2 inches in diameter, recessed in the triangle target 6 inches from the corners and in the center of the overhead target; a fencing foil wired so that contact with the copper discs activates an electric counter.

Application: This apparatus may be used to measure co-ordination. The subject stands in such a position that a lunging movement will enable him to touch any of the three discs of the triangular target, and in such a position that he may reach overhead and touch the disc of the overhead target. At the starting signal, the subject attempts to strike the copper discs of the triangular target, progressing clockwise, and then strikes the disc in the overhead target. Time required to complete 50 thrusts is recorded in tenths of a second by means of a stop watch.

Sources: Collins, Vivian D., and Howe, Eugene C. "Preliminary Selection of Tests of Fitness." *American Physical Education Review* 29:563-70; December 1924.
Masley, John W., and others. "Weight Training in Relation to Strength, Speed, and Co-ordination." *Research Quarterly* 24:308-15; October 1953.

2. Eye Blink Measurement

Basic Components: a trigger switch consisting of a small mercury trough attached to the frame of a pair of spectacles, electronic timer, light stimulus, standard electric clock, reaction time key, buzzer, and adjustable chin rest.

Application: This apparatus may be used to measure the duration of the black-out period that occurs when a person blinks his eyes. A fine piece of copper wire is attached to the eyelid so that when the eye closes the end of the copper wire dips into a mercury trough. This action causes an electric circuit to close and triggers the electronic timer. The electronic timer produces a square wave pulse which may be varied in duration in 10 equal steps from .01 to .10 seconds. When the contact is made between the copper wire and the mercury trough, the light stimulus circuit and the circuit for the standard electric clock are closed. It is possible to determine whether the subject can see light stimuli of varying durations, thereby determining the interval of the eye blink.

Source: Slater-Hammel, A. T. "Blackout Interval During Eye Blinks." *Research Quarterly* 24:362-67; October 1953.

3. Simulated Smoking Device

Basic Components: 660-watt electric heater element, asbestos covered coffee can, pyrex tubing, rubber tubing, wooden mouthpiece, blindfold, and nose clip.

Application: "Placebo" to be used in experiments relative to effects of cigarette smoking.

Source: Parker, Paul A. "Acute Effects of Smoking on Physical Endurance and Resting Circulation." *Research Quarterly* 25:210-17; May 1954.

4. Visual Acuity

Basic Components: three dimensional tachistoscope.

Application: The tachistoscope may be used in testing memory, vision, and other factors associated with vision by throwing images of objects on a screen for a measured period of time. This apparatus, which was developed by the Three Dimension Company, and other devices such as the telebinocular, orthorater, and sight screener are useful in testing the many factors associated with vision. There has not been a great deal of research done in respect to physical performance and vision, however, because of the difficulty in devising studies to demonstrate the association between these two variables.

The tachistoscope is used extensively in training people to be better readers and, no doubt, has some effect on the perception of visual images so far as the time element is concerned.

Source: Sloane, A. E., and others. "Effect of a Simple Group Training Method Upon Myopia and Visual Acuity." *Research Quarterly* 19:111-17; May 1948.

PRINCIPLES FOR PLANNING

The following principles to be observed in the planning of a research laboratory were drawn up by Raymond A. Weiss of New York University and Frank D. Sills of the State University of Iowa for the Research Council of the American Association for Health, Physical Education, and Recreation.

Planning and Administration

1. When a laboratory is planned, the planning committee should obtain from the college or university any formulated or anticipated long-range plans for growth or changes in enrollment, and find out what future emphasis in the sciences is expected.
2. The planning committee should know whether funds for the laboratory will be provided through the school budget.

Laboratory Facilities and Their Arrangement

1. Where space is limited, the laboratory should be planned to allow flexible use of available space by providing:
   a. Movable wall partitions
   b. Movable or multiple services such as water, air, electricity, etc.
   c. Portable equipment
   d. Portable furniture.
2. Where space is adequate, special tables should be provided to hold delicate instruments which should be permanently located.
3. All services such as pipes, ducts, and conduits should be concealed (in walls, floors, or ceiling).
4. Comfort of laboratory personnel and subjects should be a consideration in the arrangement of laboratory facilities.
5. Standard furniture is to be preferred over custom items because of cost.
6. Table tops or working surfaces should be specified with their most common use in mind.
7. Cabinet space should be provided to avoid using work space for storage purposes.

8. Hoods should be provided wherever smoke and fumes are generated in the laboratory.
9. Disposal sinks should be provided when chemicals are used in the laboratory.
10. All disposal units such as sinks and vacuums should be properly equipped with traps to permit recovery of materials such as mercury, grease, etc.
11. Special rooms should be provided for laboratory functions that would interfere with the use of the main laboratory space. Examples are dark room, X-ray room, and conference room.
12. Compressed air, vacuum, and gas services should be installed as standard pipe systems with outlets at tables where these services are used.
13. Full consideration should be given to the electrical requirements of the laboratory:
    a. Type of current (A.C. or D.C.)
    b. Electric distribution requirements
    c. Outlets wherever needed
    d. Circuit breakers to protect feeders
    e. Transformers where needed
    f. Installation of branch circuit panel boards where central control is required
    g. Emergency light power.
14. Artificial lighting should be planned to provide the quality and intensity of light needed for specialized functions in various parts of the laboratory.
15. Rest rooms should be provided for men and women, with separate rooms for experimental subjects and staff.
16. Dressing room and shower facilities should be provided for subjects who participate in experiments.
17. The laboratory should have its own separate stock room.
18. Office space for laboratory personnel should be provided separate from, but adjacent to, the main laboratory space.
19. Shelf space should be provided to hold reference and research materials, books, and periodicals. If space permits, table and chair facilities should be provided nearby for study and reference reading.
20. If animal research is carried on, it would be desirable to have a small animal room located close to the main laboratory facility.
21. In laboratory procedures involving the use of quantities of mercury, special tables having raised edges should be used to minimize the loss of this element.

Laboratory Equipment

1. Equipment needs should be anticipated insofar as possible and purchases made when future use is assured.
2. When possible, equipment should be purchased for permanent assignment to the laboratory. However, in the case of more expensive

equipment used for highly specific purposes, the research laboratory and other science departments in the college or university might purchase and use equipment jointly.
3. While planning a laboratory, the planning committee should visit a number of similar type laboratories to gain ideas and information that may be applied to the new installation.
4. The laboratory director should retain responsibility for planning the laboratory, its purpose and activities.
5. Wherever possible, a service engineer should be consulted in planning the equipment and facility requirements.
6. Although the laboratory may be assigned for partial use in connection with university courses, such use should be avoided if it interferes with schedules of student and faculty research.
7. In the original planning of a laboratory, provision should be made to allow for the expansion of facilities that may be needed later on.

Interior Construction Materials

1. The factor of dust—its type and amount—should be considered in the selection of materials.
2. Corrosion caused by chemicals and gases should be considered in the selection of materials.
3. Noise inside or outside the laboratory should be considered in the selection of materials.
4. Vibration inside or outside the laboratory should be considered in the selection of materials.
5. The addition of moisture in the air as a result of research should be considered in the selection of materials.
6. Comfort should be considered in the selection of materials (cork floor to reduce fatigue in standing, etc.).
7. The control of light admission should be considered in the selection of window assemblies.
8. The control of ventilation should be considered in the selection of window assemblies.
9. The size of door and door opening should be adequate for admitting laboratory facilities and equipment.

Care and Maintenance

1. All trim and molding should be eliminated in the laboratory to facilitate cleaning.
2. The laboratory should have a self-contained workshop to provide maintenance for equipment and facilities and to construct equipment that is otherwise not available.

Safety and First Aid

1. Separate areas should be provided for inflammable liquids, compressed gases and chemicals.

2. Nonslip finishes should be provided for floor surfaces to prevent slipperiness under wet conditions.
3. Moving parts of all mechanical equipment should be fenced off from contact by means of handrails with toeboards or should be enclosed within stationary guards.
4. There should be protection against fire hazard through provision of chemical fire extinguishers, woolen blankets, or emergency showers.
5. Medical personnel should supervise any research which imposes severe physical stress on subjects or which requires injections or blood withdrawals for biochemical analyses.
6. All personnel working in the laboratory should be required to take out liability insurance if such a policy is not provided by the institution.
7. A well-stocked first aid cabinet should be located for quick and easy access in case of emergency.

CHAPTER 7

Tools for Analyzing and Presenting Data

H. HARRISON CLARKE
MARJORIE PHILLIPS

THE DISCUSSION OF TOOLS FOR ANALYZING AND PRESENTING DATA will be restricted primarily to definitions, limitations, underlying assumptions, and uses of the various statistical concepts. For the process of making statistical computations, the researcher should consult standard statistics textbooks.

NATURE OF STATISTICS

In considering the meaning of statistics, Helen Walker (19:1) has provided the following concept:

> "Statistical method is one of the devices by which men try to understand the generality of life. Out of the welter of single events, human beings seek endlessly for general trends; out of the vast and confusing variety of individual characters, they continually search for underlying group characters, for some picture of the group to which the individual belongs."

Statistics deals with the collection, the organization, the analysis, and the interpretation of quantitative data—e.g., the test scores of many individuals. The use of statistics is found in many fields devoted to research, including agriculture, economics, government, sociology, biology, psychology, medicine, and physical education, to name several. Statistics permits treatment of data in these fields, making possible study of many problems peculiar to them.

Scientific research involving measurement results in quantitative data. Statistical methods need to be applied in order to gain a summarized description or analysis of the findings. In order to accomplish these purposes, it is essential to understand first, what kind of description is wanted; second, what statistic will yield the most valid description; and third, whether the assumptions underlying the selected statistic are satisfied by the data being described.

Three classes of statistical processes will be considered in this chapter: descriptive statistics, comparative statistics, and statistics of inference. In descriptive statistics, the characteristics of a single group are described in various ways. In comparative statistics, the characteristics of two or more groups are contrasted. In statistics of inference, observed data from a sample are used as a basis for generalizing to a larger unknown population which has not been observed.

MEANING OF DATA

In general there are two types of statistical data familiar to the areas of health education, physical education, and recreation; these are *attributes* and *variables*. An attribute has a nongradient classification, i.e., there is no numerical basis of grouping. Attributes may be in two classes or more than two classes. Examples of two-class attributes are pupils as boys or girls, teachers as men or women, and curriculums as college preparatory or noncollege preparatory; more than two-class attributes are color of hair or eyes, various curriculums, and major fields of study.

A variable has a gradient classification, i.e., there is a numerical basis of grouping. There are two types of variables—continuous and discrete. A continuous variable is capable of any degree of subdivision, although in practice these are usually limited to some convenient number of divisions. Most of the test data used in physical education research fall in this category. Illustrations of continuous variables are muscular strength, anthropometric measures, track and field times and distances, and motor ability scores.

A discrete variable cannot be, or is not generally, subdivided indefinitely. Illustrations are football scores, salary scales, buildings, and numbers of pupils in a classroom. While fractions of such scores are not realistic (Who ever heard of a softball score of 5.48 points?), discrete data are frequently treated statistically as though they were continuous. Certain comparisons would not be possible otherwise. For example, to state that the average number of children in families from two different economic groups are 2.63 and 3.42 is to state the impossible; actually, however, no other comparison would be adequate, as to round off the figures to the nearest whole numbers would result in three children for each group.

FREQUENCY DISTRIBUTION

The first task in dealing with a large number of test scores is to organize the data. This is accomplished by assembling the scores into groups or classes, thus constructing a frequency distribution table. The frequency table consists of intervals, each of the same size, and includes the entire range of scores. For example, a series of scores from 75 to 125 might be placed into ten intervals, in units of five each, beginning at 75.

There are two assumptions that are usually made when calculations are made from a *grouped* frequency distribution, as follows:

1. The scores are evenly distributed within the interval. This assumption is made when computing percentiles and percentile ranks, and is made only in the interval being interpolated. This particular assumption is necessary because the process of interpolation demands a linear scale. Error will be limited to the one interval, and relative status will not be disturbed.

2. The mean of the scores within the interval is equal to the midpoint of the interval. This assumption is made for every interval in the distribution and is used when finding the mean of the distribution, or any measure based on the mean. The midpoint is selected to represent all the scores in the interval, since it is the one point in the interval from which there will be no final error resulting when the distribution is bell-shaped. The assumption becomes necessary since each score must have an identity, if the arithmetic calculation is to be made.

A slight error because of the above assumption may result as a consequence of the natural tendency of scores in the intervals to crowd toward the center of the distribution. The averages of scores in the upper intervals of the distribution tend to be slightly lower than their midpoints; and the averages for the lower intervals tend to be slightly higher. Since the interval errors tend to cancel out, the amount of final error is so small as to be generally disregarded.

CENTRAL TENDENCY

A measure of central tendency is a single score or point on a scale which represents all the scores made by a group. If one is asked how well pupils performed on a particular test, the answer will be that the average was so-much. Thus, in common parlance, a measure of central tendency was used. The measures most frequently encountered in educational research are the mode, median, and mean. Each measure of central tendency describes the massing of scores in a particular manner.

Mode. The mode is the point in the distribution at which the largest concentration of scores occurs. When scores are ungrouped, the mode is the score most frequently occurring. With grouped data, the mode is designated as the midpoint of the interval containing the largest number of scores; when adjacent intervals have equal or nearly equal frequencies, the midpoint of the combined intervals becomes the mode. It is also possible to compute an approximation of the mode from the median and mean of the distribution.

The mode is used when information is desired as to a phenomenon which happens most frequently. For example, when teaching skills, the skill which is most difficult to learn would be the mode, as the largest number of participants have trouble in learning it. In some situations, the mode is the desired measure for designating central tendency. An example would be the age of children in a single grade in school or the age of high school graduates. Also, it is the only measure of central tendency which can properly be used to describe the typical performances of a multimodal distribution. In general, however, the mode is usually a rough measure of central tendency and has little value in exact statistical work.

Median. The median is the midpoint of the distribution: that point above which and below which lie 50 percent of the scores. When scores are ungrouped, but arranged in order from high to low, it is quite easy to find the center, or middle, score. With scores grouped in a frequency table, the median is found by counting frequencies to the interval in which the middle score lies, and then interpolating to determine the point. In this instance, the assumption is made that the scores in the interval of calculation are equally distributed over the interval. No assumption is made relative to the nature of the entire distribution.

There are several situations in which the median provides the best indication of typical performance. The most common of these is the presence of extreme scores at one end of the distribution. The actual size of scores within the distribution does not affect the median, as the scores are merely counted in making the calculation. The median is also the best measure when the distribution is truncated. A truncated distribution would occur, for example, as a result of using a strength-testing instrument with a capacity less than the strongest subjects. Furthermore, the median is preferred when the equality of the unit of measurement is uncertain, as in a rank order of performance, and when numerical measurement is impossible, as in an arrangement of silhouettes depicting poor to good posture.

Mean. The mean is the sum of the scores divided by their number; thus, each separate score affects this measure of central tendency in direct proportion to its magnitude and position in the distribution. The mean expresses the central massing of scores according to the distance the scores fall from the mean. It is, therefore, a deviation measure of central tendency, as each score in the distribution is weighted by its distance from central tendency. With upgrouped data, the mean is calculated by adding the scores and dividing by the number. With grouped data, it is obtained from the sum of the positive and negative deviations of the scores from an assumed central position (midpoint) in the distribution. Thus, a basic assumption in using the mean, when calculated from grouped data, is that the scores in the various intervals composing the frequency table are represented by their respective midpoints, or that the mean of the scores in a given interval is equal to the midpoint of the interval.

The mean is the most reliable of the measures of central tendency, i.e., there is less fluctuation among sampling means than is true for the mode and median. The mean should be used when distributions are reasonably symmetrical, when they are not skewed, or when they do not contain extremes at one end of the distribution. Actually, it should always be utilized as the measure of central tendency unless the mode or median is more appropriate, as indicated above.

VARIABILITY

In most instances, it is not only necessary to describe the central massing of scores, but also the amount of their variability. The mean score may be the same value in a distribution with wide variability as in one with narrow variability. Thus, measures of variability need to be considered. Such measures indicate the scatter or spread of the various scores, usually around a measure of central tendency. While central tendency is a point on a scale (within a distribution), variability denotes distance on a scale. The variability measures to be presented here are range, quartile deviation, average deviation, standard deviation, and probable error. The coefficient of variability will also be briefly described.

Range. The range indicates the scatter or spread of all the scores in the distribution. It is obtained by finding the difference between the highest and lowest scores, without reference to measures of central tendency. In chance sampling, the range is subject to greater fluctuations than any other measure of variability. It is used when knowledge of the extreme scores of human characteristics is desired. Thus, the range of the height of basketball players or the weight of football players might not only be interesting but be valuable.

Quartile Deviation. The quartile deviation indicates the scatter or spread of the middle 50 percent of the scores taken from the median. Also known as the semi-interquartile range, it is calculated by dividing the difference between the third and first quartiles by two. If the distribution is fairly symmetrical, this distance added to and subtracted from the median approximates these quartiles. This measure of variability is usually used when it is appropriate to use the median as the measure of central tendency.

Mean Deviation. If the distribution conforms to a normal curve, the mean deviation indicates the scatter or spread of the middle 57.5 percent of the scores taken from any measure of central tendency. It may be preferred to the standard deviation when the distribution is markedly skewed or has an unusual number of extreme cases in one or both directions from the mean. However, this measure of variability is infrequently found in research, as its calculation does not conform to proper algebraic procedure.

Standard Deviation. The standard deviation, or sigma, indicates the scatter or spread of the middle 68.26 percent of the scores taken from the mean. It is calculated as the square root of the average of the squared deviations from the mean. The assumption underlying sigma is the same as the one for the mean. The use of standard deviation assumes a normal distribution if interpretations are made relative to a normal distribution, as is frequently the case.

The standard deviation is usually utilized when the mean is the appropriate measure of central tendency. As is true with the mean, this measure of variability is affected by the size and position of the separate scores in the distribution. The standard deviation is the measure of variability customarily used in the statistical analyses basic to experimental research.

Probable Error. If the distribution is normal, probable error indicates the scatter or spread of the middle 50 percent of the scores taken from the mean. It is calculated from the standard deviation, so it has all the characteristics of this measure of variability. Probable error was used extensively in earlier statistical analyses, so it appears frequently in the older literature. However, its use in current research is very limited.

Coefficient of Variation. The coefficient of variation provides an indication of relative variance of groups. It is useful when the central tendency of groups on the same test differs or when results of different tests are to be compared. It is calculated as the ratio of a measure of variability to its appropriate measure of central tendency. Thus, this coefficient may be calculated with sigma and mean or with quartile deviation and median. Such questions as the following might be answered: Are college men more variable in weight than in height? Are 12-year-old boys more variable in reaction time than girls of the same age? Do varsity swimmers vary more than nonswimmers in breath-holding?

An essential assumption in using the coefficient of variation is that a true zero point exists. For example, if plus and negative values exist, the mean could be zero, in which case the variability coefficient would be zero. The true zero point should be known in a testing situation, which is not always the case in education and psychology. Thus, V is not a thoroughly reliable measure by means of which to compare the variability of two distributions. However, no better measure for this purpose has been proposed, so it continues in use.

NORMAL PROBABILITY CURVE

An understanding of the characteristics of the normal probability curve is essential to an understanding of reliability, that important phase of statistics dealing with the interpretation of statistical results. It is only through measures of reliability that the true usefulness of such obtained measures as means, standard deviations, and coefficients of correlation can be understood.

Principles. The normal curve is based upon the probable occurrence of an event when that probability depends on chance. For example, in flipping a coin the chances are even, or one in two, that it will come down heads, and there is the same probability that it will come down tails. This chance probability can be carried to flipping many coins, in which case an exact theoretical explanation of the curve is provided by the binomial theorem. This can be demonstrated by expanding and plotting the equation $(H + T)^n$.

This theory of normal distribution, as applied to the chance occurrence of heads and tails in coin tossing, is also applied to the chance occurrence of many human characteristics. Thus, anthropometrical, biological, psychological, and social traits cluster about an average and will be distributed in much the same way as are the heads and tails in coin tossing. The occurrence of the normal curve, however, whether in coin tossing or in the occurrence of human attributes, depends upon two very important factors: (a) The occurrence of the event must depend upon chance, i.e., the sample must be selected at random; (b) A large number of observations must be made.

Characteristics. The normal curve is a symmetrical, bell-shaped curve. Its characteristics are: (a) It is asymptotic to the baseline; (b) The points of inflection are each one standard deviation from the ordinate at the mean; and (c) The height of an ordinate at any given standard deviation distance from the mean ordinate is an exact proportion of the height of the mean ordinate. Thus, the area under the curve included between the mean ordinate and an ordinate at any given standard deviation distance from the mean will be an exact proportion of the total area under the curve.

The most important factor related to the normal curve is the division of the curve into percentage areas. Knowing the mean and standard deviation and knowing that the distribution is normal, it is possible to obtain from a standard table the percentage of scores falling between the mean and any given standard deviation distance above or below the mean. Thus, from a standard table, 34.13 percent of the scores are between the mean and one sigma away from the mean in the same direction. Nearly one-half the cases, 49.865 percent, lie between the mean and three sigmas; thus, for plus and minus three sigmas, 99.73 percent of the cases are included.

Testing for Normality. Because of the value and utility of the normal curve, it is frequently desirable to test a given distribution for normality. Reasons for deviations from normal include the presence of some bias in the sample; the use of unsuitable tests; the selection of small homogeneous or large heterogeneous groups; and the fact that some traits do not distribute normally (e.g., the curve of forgetting). Slight departures from normality will not necessarily influence the accuracy of the values derived from the curve, but tests to determine the significance of any departure should be applied.

There are two methods commonly used for the purpose of testing the significance of distortion from the normal curve found in a frequency distribution. These are the inspectional and chi-square tests. The *inspectional test* simply consists of superimposing a frequency polygon on a theoretically calculated normal curve for the mean and sigma of the distribution to be tested. A fault with this method is that it provides no tests to determine the significance of departure from normality. In the *chi-square test,* as will be considered below, the actual frequencies in each interval of an obtained frequency table are compared with the theoretical frequencies needed to yield a normal curve. A test of significance, in this instance, does determine the importance of deviations from normal. This test, however, does not indicate the ways by which a given distribution departs from normality.

In addition to the general determination of normality, other tests are available for phases of normality. The most common of these are *skewness* and *kurtosis*. A distribution is said to be skewed when the mass of data is shifted to the upper or lower end of the scale. In positive skewness, the scores are massed at the lower end with a wide spread of scores at the upper end; the reverse is negatively skewed. The calculated skewness value should be tested to determine its significance, that is, whether it is real or due to chance sampling fluctuation.

Kurtosis indicates the nature of the distribution at the center. A peaked curve, i.e., higher than the normal curve at the center, is called *leptokurtic*. A curve flatter at the center than normal is called *platykurtic*. A normal curve is *mesokurtic*. Any deviation from a mesokurtic distribution needs to be tested to determine the significance of the distortion. This test will show whether the departure from normality can or cannot be attributed to chance and the sample was or was not drawn from a population which is normal in form.

SCORING SCALES

The use of scoring scales is extremely valuable in making test scores meaningful and in describing the results achieved by a class or group of individuals that has been tested. It is impossible to know how well one has done on a test unless his score is shown in relationship to others taking the same test. Also, with scoring scales, the individual's relative performance in events with quite different forms of measurement can be determined. How else, for example, can performance in the standing broad jump be compared with the time in the 50-yard dash? Usually, these scales are based on a 100-point range, although there are exceptions to this.

Percentile Scale. The percentile scale is based on percentages below points on the scale. Thus, the median is the 50th percentile, as 50 percent of the scores fall below this point. In like manner, 10 percent of the scores are below the 10th percentile, 75 percent below the 75th percentile, and so on. Since percentiles are located by a counting process and are nonarithmetical in nature, further calculations should not be based upon them. Thus, their use for additional research based on a measure of relative status is limited.

Sigma Scales. Sigma scales are based on the variability of the group, using the standard deviation as the measure of variability. These scales are based on sigma divisions of the distribution. The following scales will be considered: Z-score, T-scale, H-scale, and six-sigma scale.

The standard score, or Z-score, is designated as the standard deviation distance of a score from its mean. The mean score itself is 0 and the standard deviation 1; thus the scale range is approximately plus and minus 3 sigmas. Hence, a Z-score of + 1.5 is a score located 1.5 sigmas above the mean; and a Z-score of − .75 is located .75 sigma below the mean.

The other sigma scales substitute a different numerical scale for the sigma distances. For these scales, the mean is usually 50, but the standard deviations differ as follows: 10 for the T-scale, 14 for the H-scale, and 17 for the six-sigma scale. Thus, in the

T-scale, 0 is located five sigmas below the mean and 100 is located at five sigmas above the mean. For the other scales, these locations are 3.5 and 3.0 sigmas respectively. An obvious circumstance with the T-scale is that it extends well beyond the limits of probable scores for normal distributions; the scale values below 15 and above 85 are seldom found. The other scales are more applicable to realistic testing situations, although scores may rarely be encountered which fall outside their limits. Actually, there is no organic difference in the various sigma scales. It is simply a matter of the range of scores which one wishes to use.

SAMPLES AND STATISTICS

Sampling is of inestimable importance to the researcher who wishes to discover information about populations when it is either unfeasible or impossible to measure an entire population. Sampling is a common procedure in all areas of inquiry today. It is used in such diverse fields as business, education, science, agriculture, medicine, and public opinion. Much useful information can be inferred, for example, from the observations secured from a few experimental animals, a handful of corn, a small section of human tissue, or a group of school children.

Sampling Theory. If the inferences concerning facts or characteristics of a population are to have validity, it is obviously of fundamental importance that the sample be truly representative of this population. It also should be obvious that any measure obtained from a sample (statistic) may differ somewhat from the true measure (parameter) in the population. The amazing truth is, however, that if proper procedures are followed, it becomes possible to estimate the amount of this sampling error, and to provide an expression of the degree of confidence that can be placed in the estimate.

Basic to sampling theory is the concept of the random sample, since it is only for such samples that the probable divergence of the statistic from the parameter may be estimated. A sample is considered to be random when all its members are selected independently and in a manner which guarantees every member of the population an equal chance to be selected. A sample is biased, on the other hand, when its members are selected in a manner which, in repeated samplings, will produce a systematic sampling error.

An understanding of the elementary principles of sampling and estimation of error may perhaps be gained through a simple illustration. The physical condition of the young men in a community, state, or country is a matter of primary concern to many people. Let us suppose that a random sample of 200 of the 18-year-old boys in a large community is selected and that a valid test of fitness is available which is administered to each of these boys. From the measures thus made available, the mean fitness of the boys is found to be a value of 75. This sample mean may not be accepted unconditionally but should be considered only as an approximation of the true value in the population. In other words, the true mean physical fitness of all the 18-year-old boys in the community may well be some other value than that of the sample mean. However, since the sample is a part of the population, there are limitations on how much the statistic and parameter may differ from each other. Furthermore, since the members of the sample were drawn at random, chance will dictate the size of the sampling error in the mean selected, or the difference between the sample mean and the population mean. The methods available for estimating the size of such sampling errors are based on the sampling distribution of the statistic, in this case the mean.

Large Samples. Let us suppose that a very large number of samples, of 200 cases each, was drawn as previously described and from the same population of 18-year-old boys. It would be observed that the sample fitness means would have a variety of values depending on the physical condition of the boys who, by chance, happened to be selected in each sample. It would be noted further, that while these means would be distributed over a considerable range of scores, a large proportion of them would tend to be grouped around a central value, and only a few means would have extreme deviations from this middle value. Such a distribution is called a sampling distribution of means, and it has been demonstrated that distributions of this kind have certain characteristics. If the samples are large, and each sample is drawn independently at random, the distribution of means will closely approximate a normal distribution—even in cases where the measures in the population are not normally distributed (except where extreme departures from normality occur). Furthermore, the mean of this sampling distribution of means would have the same value as the true mean in the population from which the samples were selected.

Under these conditions, it should be possible to visualize how an estimate of reliability for any sample mean is derived. Referring to the physical fitness problem, the sample mean of 75 should be thought of as having been drawn at random from a sampling distribution of means, the mean of this distribution being the true physical fitness mean of the entire population of boys in the community. Since the sampling distribution is normally distributed, it becomes evident that it is now possible to state the probability that the sample mean of 75 is any given number of standard deviation units from the true mean. For example, there are only 4.56 chances in 100 that the sample mean will be more than two standard deviations from the true mean, or .26 chances in 100 that it will deviate more than three standard deviations from the true mean. If the value of the standard deviation of the sampling distribution were known, the amount of separation between the sample mean and the population mean could be expressed in the score units of the physical fitness test.

In sampling theory, the standard deviation of the sampling distribution of a statistic is known as the standard error of the statistic. The standard error of a mean has been found to be dependent on two factors—the size of the sample from which the mean was derived and the variability of the population from which the sample was drawn. From these facts it has been possible to develop an unbiased estimate of the standard error of the mean. Continuing with the physical fitness problem, assume that the standard error of the mean has been found to have a value of 2. A more exact statement may now be made about the probable deviation of the sample mean of 75 from the population mean. It has already been shown that there are only .26 chances in 100 that the sample mean will be more than three standard deviations from the true mean. It therefore follows, since the standard deviation of the sampling distribution (standard error of the mean) has a value of 2, that there are only .26 chances in 100 that the sample mean will be more than 6 points (3 x 2) from the population mean. In this fashion, the standard error of the mean becomes a measure of the reliability of any mean in the theoretical sampling distribution. The larger the standard error, the less reliable is any sample mean; and the smaller the standard error, the more reliable is any sample mean. This same rationale may be employed in estimating the reliability of any statistic—such as the median, quartile deviation, standard deviation, proportion, difference, correlation, and so on.

The procedures just described may be used to estimate that part of sampling error which arises from the random selection of the members of the sample. If any other source of error is present or if the sample is not a random one, the reliability estimate is invalidated, since it evaluates chance errors only. The burden is therefore placed on the investigator to exercise the utmost caution in the preparation of his study. It is his responsibility to eliminate avoidable errors since they will remain unmeasured, and to draw the members of his sample at random to permit the measurement of the unavoidable errors arising from this source.

Levels of Confidence. Before discussing specific applications of sample theory, the meaning of a "level of confidence" should be explained. The purpose of studying samples is to draw conclusions relative to populations. However, conclusions based on random samples may never be stated without qualification, because of the part chance plays in reaching the conclusion. Since the effect of chance is never completely eliminated, it is proper procedure, in certain instances, to reveal the probability that the conclusion drawn is incorrect. In the physical fitness problem, for example, it is known that there are only .26 chances in 100 that the sample mean of 75 is more than 6 points from the population mean. Stated in a different way, it may be concluded that the sample mean of 75 deviates no more than 6 points from the population mean, at the .26 percent level of confidence. The level of confidence thus expresses the probability that the sample mean actually deviates more than 6 points from the population mean and that the conclusion drawn is incorrect.

A second illustration may help to clarify the essential meaning of a level of confidence. An individual has a bag of marbles containing 190 white marbles and 10 red marbles. The marbles are thoroughly mixed, and he selects one marble strictly at random, without being able to see the contents of the bag. Previous to his draw he announces, "I will draw a white marble." It is obvious that the probability is high that he will draw a white marble, but since 5 percent of the marbles are red, there are actually 5 chances in 100 that the single marble drawn will be a red one. Hence, to be absolutely correct in his statement, the individual should announce, "I will draw a white marble, at the 5 percent level of confidence."

Confidence Intervals. *It has been demonstrated, in the fitness problem, that the sample mean of 75 will be no more than 6 points from the parametric mean, at the .26 percent level of confidence.* It must follow then that the value of the true mean cannot be lower than 69 or higher than 81, at the .26 percent level of confidence. In other words, the mean physical fitness of the population of 18-year old boys, whatever value it may actually be, will not be less than 69 or more than 81. The chances that these limits will be exceeded are only .26 in 100.

Expressing the limits within which a parameter may be expected to be contained is called a confidence interval. In the illustration above, the .26 percent confidence interval was given for the true mean. The confidence interval may be given in the same manner for any other parameter, such as the true standard deviation, true difference, or true correlation. All that is needed is the statistic and its standard error. The confidence interval may also be given for any level of confidence that suits the purpose of the investigator, although rarely, if ever, would a level of confidence below 5 percent (a larger numerical value) be selected. The levels of confidence most frequently used are the 5 percent and the 1 percent levels.

Testing Hypotheses. In a sampling problem, the value of the parameter is never known; if it were, there obviously would be no need to sample. It is possible, however, to propose some hypothesis about a parameter and then test its plausibility. Since the procedure of testing hypotheses provides a basis for experimental research and statistical inference, an understanding of it is of vital importance.

A simple illustration of testing a hypothesis is given by the fitness problem. Let us assume it is known that for 18-year-old boys throughout the country the mean physical fitness is 78. The mean physical fitness in the community as found from the sample is only 75. Does this provide evidence that the boys in the community are below the national level in fitness? Or, is it possible that this sample mean of 75 was drawn from a population in which the true mean is 78, the observed discrepancy being the result of the chance selection of the sample members?

To answer these questions, the following hypothesis is tested: In the community population from which this sample mean of 75 was drawn, the true mean physical fitness of the 18-year-old boys is 78. If this hypothesis is true, then the sample mean of 75 contains a sampling error of 3 (78-75). The standard error of the mean is known to be 2, hence the size of the sampling error expressed in standard deviation units is 1.5 (3/2), and is referred to as the critical ratio. It now becomes a simple matter to determine the chances of having drawn a sampling error of 3 or larger. In any normal distribution, 1.5 standard deviations will be exceeded by 13.36 percent of the cases; hence in this problem, there are 13.36 chances in 100 of having drawn a sampling error of 3 or larger.

At this point, the decision must be made whether to accept or reject the hypothesis. The hypothesis is accepted if it may be considered reasonable to suppose that a sampling error was drawn that, in the long run, would be drawn 13.36 percent of the time. The hypothesis would be rejected if this supposition were considered unreasonable. When a hypothesis is accepted, it does not mean that this particular hypothesis is necessarily true—only that it is reasonable to believe that it could be true. When a hypothesis is rejected, it means that, from the evidence available, the fact in the sample and the hypothetical condition in the population are incompatible. Since the sample has been under actual observation, it must be the hypothetical condition which is in error. However, since even the remotest chances materialize on occasion, the rejection of the hypothesis is always qualified by the level of confidence. In the illustrative problem under discussion, the hypothesis would be accepted. The investigator concludes that it is a reasonable supposition that the true mean physical fitness of the community of boys is 78. The evidence does not support the contention that the boys in the community are below the national fitness level. Had the hypothesis been rejected, the level of confidence would have been 13.60 percent, hardly a sufficiently high level to inspire confidence in a rejection.

In the solution of any problem for which the critical ratio technique is used, all that is needed is the size of the hypothetical sampling error and the standard error of the statistic, since the critical ratio is the ratio between these two. This technique has many applications and has been found particularly useful in testing hypotheses concerning means, proportions or percents, and differences between means.

Let us suppose that the health education authorities in a community are faced with the problem of introducing sex education into the public schools. Before making any overt moves, they wish to be assured that at least a majority of the parents will support such instruction. A large random sample of the parents of the children in the school is drawn, and these parents are interviewed relative to their attitudes towards sex education in the public schools. The survey reveals that 58 percent of the parents favor such instruction. Does this constitute evidence that a majority of the population of parents approve sex education, or is it probable that this sample percentage of 58 came from a population in which less than 50 percent of the individuals favor sex education? The appropriate hypothesis to test in this situation is that the true percent of parents favoring sex education is 50. If this hypothesis may be rejected at a reasonable level of confidence, then any hypothesis representing less than a majority may be rejected with an even greater assurance.

A second illustration is drawn from the field of public health. In this case, the officials in a country are concerned about the incidence of a nutritional deficiency disease, such as beriberi or scurvy, among children in poverty-striken areas. A random sample of the children reveals that a certain proportion is affected. An estimate of the true incidence may be given through the confidence interval (A special procedure is necessary for computing the confidence interval for a proportion or percent.), or any hypothesis may be tested concerning the true proportion of children who have the disease. If the normal incidence of the disease is known for the country, that might well be the hypothesis selected.

One of the most useful of all statistics is a difference, since it permits a comparison of results obtained from two random samples. While the difference between two means is frequently the point of primary interest and is therefore the difference most often tested, the procedures are equally applicable to other statistics, such as medians, standard deviations, correlations, and so on.

Let us assume that two physical education teachers in neighboring communities give the same strength test to a random sample of children from their respective elementary school populations. The mean strength score for 10-year-olds in Community A is 110 and in Community B is 115—a difference of 5. May this difference be attributed to the chance error arising from the random selection of the sample members, or is it evidence of a real superiority in strength for the children of Community B? The hypothesis to be tested is the null hypothesis, which states that the true difference is zero. If this hypothesis may be rejected, at a satisfactory level of confidence, the difference is regarded as significant, that is, too large to be attributed reasonably to chance. In this case, the evidence supports the superiority of Community B children over Community A children in strength. If, on the other hand, the hypothesis is acceptable, the evidence indicates that the true difference may well be zero and that in terms of strength the children of the two communities are regarded as being from a common population.

The testing of differences has numerous applications in learning and training situations. Let us consider the case in which a swimming instructor wishes to evaluate the comparative effectiveness of two types of training for improving flutter-kick speed. Two random groups are selected and the groups are trained by the two different treatments under consideration. At the end of the training period, the subjects are measured and the mean speed for some specified distance is found for each group. The null hypothesis for the difference between the mean speeds of the groups is then tested. If the null hypothesis is accepted, it indicates that there is good reason to believe that the true difference in mean speed of the two groups is zero, or that, under the conditions of the investigation, the evidence supports the idea that the two types of training are equally effective in developing flutter-kick speed. If the hypothesis is rejected, the conclusion that a real or significant difference exists is considered reasonable; or one of the groups is superior in speed to the other, the distinction being made by examining the relative size of the mean speeds. The conclusion of a significant difference does not, in itself, constitute evidence as to the cause of the difference. The statistical analysis simply indicates that whatever the cause of the difference, it is too large to attribute all of it to chance. To reach the conclusion that one of the types of training is more effective than the other, the investigator must show that all factors which conceivably could have caused such a difference were constant for the two groups, with only the two types of training being permitted to vary.

The example just given illustrates a study in which two independent groups were observed. There are some cases where a better solution to a problem may be found by using "related groups," that is, groups in which the members are matched or paired. When a relationship exists, it must be given consideration in the standard error of the difference formula, and the planning and execution of the investigation will be affected by this factor. (See Chapter 10 on the Experimental Method.)

Small Samples. The rationale described for large samples is equally applicable to small samples. It has been found, however, that sampling distributions developed from small samples are not normally distributed. These distributions vary for different sizes of samples, with the tails of the distribution becoming more extreme as the sample size decreases. Therefore, for small samples the evaluation of sampling errors may not be based on the Table of Areas Under the Normal Curve; instead, a special table, the Table of t, has been developed for this purpose. The symbol t has the same meaning for a small sample that CR (critical ratio) has for a large sample, in that both are the ratio between the sampling error and the standard error of the statistic. In any situation where doubt exists as to whether to use the Table of t or the Table of Areas Under the Normal Curve, the Table of t should be used. Examination of the two tables reveals that as the number of cases exceeds 100, the values in the two tables are quite close and eventually, as the sample size becomes very large, the values become the same. As the sample size decreases below 100, the discrepancies between the two tables become increasingly greater.

If, in a small sampling study, the primary interest is the significance of the difference between two means, there is a factor which should be given some consideration. The t test in such a case is a test of the hypothesis that the samples were randomly drawn from a common normal population. If t is significant, this hypothesis may be rejected, but this does not constitute a specific test for the difference between means. It is possible, although not very probable, that the significant t was caused by a difference in the variances only. In cases where some doubt concerning the variances exists, the F test for the difference between variances may be applied. If the F test reveals a highly significant difference between the variances, introducing uncertainty as to the validity of the interpretations for the means, a different test of significance may be applied. The hypothesis tested would be that the samples were randomly drawn from different normal populations with the same mean. A nonparametric test might also be useful in a situation of this kind.

Chi Square. Chi Square (χ^2) is one of the most versatile of all the statistics available, since it may be used to test various hypotheses and is applicable to a variety of situations. The simplest applications of a chi-square test occur for those cases in which the subjects of the sample may be divided into two or more mutually exclusive categories. The observed frequencies which fall into the various categories are then tested to determine whether their divergence from theoretical frequencies is too great to have occurred by chance, or conversely, whether the divergence is so small that it could reasonably be attributed to chance. The theoretical frequencies are derived on the basis of the hypothesis being tested.

A simple illustration of this application of χ^2 is provided by the following situation. A community recreation director wishes to give consideration to the interests of the adult population in his community before developing the schedule for certain hobby groups. A random sample of 150 adults indicates that 56 prefer art, 45 music, and 49 drama. On the basis of this evidence, it might be assumed that art groups should be scheduled with greater frequency than the other two, and music groups with the least frequency. However, realizing the role that chance plays in the selection of a random sample, it might be far wiser to speculate on the possibility that chance is directly responsible for the variations in the observed frequencies and that in the population sampled there actually is no preference for the three hobby groups. If the hypothesis of no preference were tested, then the theoretical population ratio would be 1 : 1 : 1; and if the sample corresponded exactly with the facts in the population, the frequencies would have been even, with 50 persons selecting each hobby group.

The organization of the data would be as given below.

Hobby Preference	*Observed Frequency*	*Theoretical Frequency*
Art	56	50
Music	45	50
Drama	49	50

In this problem, the divergence of the observed frequencies 56:45:49 from the theoretical frequencies 50:50:50 would be tested. If it is found that this divergence may reasonably be attributed to chance, the recreation director would have a basis for scheduling equal numbers of the three hobby groups. Otherwise, the results would indicate a need for scheduling more of one kind of group than another.

This type of test is known as a test for "goodness of fit," and any hypothesis concerning the population ratio may be explored by this class of tests. One of its applications in the area of knowledge test evaluation serves to emphasize the variety of situations in which it may be useful. An individual has developed a knowledge test composed of multiple choice items, with each item having five responses. His problem is to learn whether the incorrect responses for each item are equally plausible. From an examination of the first item, it was found that 60 students in the sample answered it incorrectly. Their choices of the incorrect responses were distributed as indicated in the table below.

Response	*Observed Frequency*	*Theoretical Frequency*
1	20	15
2	10	15
3	17	15
4	13	15

Under the hypothesis of equal plausibility, the theoretical frequency for each response becomes 15 and the investigator applies the χ^2 test as a test of the compatibility of the observed and theoretical frequencies.

Another example of a test for "goodness of fit" occurs when an investigator is interested in learning whether the frequency distribution for some trait conforms to a specified distribution in the population. For example, in the fields of health and physical education, an investigator may wish to develop norms for a physical fitness test. Since the norms are to be based on the mean and standard deviation of the sample, it is desirable to have reasonable assurance that fitness as measured by this test is normally distributed in the population. The answer to this problem may be secured by applying the χ^2 test for "goodness of fit."

A second class of tests involving χ^2 is sometimes called "tests of independence" or "tests of homogeneity." In either case, the test is for the presence of relationship between two traits. This application of χ^2 is particularly valuable when the measure of the traits is qualitative, and it may be applied to either ordered or unordered categories. For example, the school health authorities in a certain large community are confronted with the problem of improving the milk consumption among the children in the elementary school population. A random sample of 200 children is selected, and the children are served a mid-morning bottle of milk. Four flavors of milk—plain, chocolate, strawberry, and orange—are rotated for 16 days and a record is kept of the reaction to the various flavors by noting the bottles consumed. The reactions are recorded as indicated below.

	Flavor Reaction		
Flavor	*Consumed*	*Partially Consumed*	*Not Consumed*
Plain	400	178	222
Chocolate	505	145	150
Strawberry	475	160	165
Orange	450	170	180

The hypothesis being tested is that there is no relationship between milk consumption and the flavor of the milk. If this hypothesis can be rejected with a reasonable degree of confidence, the presence of a relation is indicated and the school health authorities would give consideration to milk flavors in planning school lunches.

The presence or absence of relationship may thus be determined for traits with any number of categories. It is important, however, to understand that this test does not provide an index of the degree of relationship, although such an index may easily be derived from χ^2 when certain conditions are met (9:392-96). The relative size of χ^2 simply provides an expression of the confidence one may have that some relationship exists.

Several special applications of χ^2 have been developed, such as the tests for homogeneity of variance, linearity of regression, and the hypothesis that several sample r's have been drawn from a common population.

The sampling distribution of χ^2 is dependent only on the number of degrees of freedom available in the table from which the calculations are made. The size of sample is irrelevant except in the case of samples of less than 50 cases or when any theoretical frequency is less than 10 (13:34). One limitation of tests of "goodness of fit" is that the direction of the differences is given no consideration. This is most serious when the test is applied to the hypothesis that a sample distribution conforms to some specified form in the population, such as a normal distribution. An examination of the pattern of the signs of the differences would be indicated in such cases and, if necessary, some more efficient test of "goodness of fit" should be performed.

In using χ^2 in the solution of problems, it is important to understand that basic to a valid interpretation of the results is the assumption of random sampling.

CORRELATION

In the above presentation of statistical concepts, a single variable has been described in various ways. This section deals with the determination of relationship between variables. Correlational methods are particularly applicable when the amount of relationship is desired, rather than the amount of change from one situation to another. A coefficient of correlation is a single measure that tells the extent to which things are related, such as human traits of various kinds present in the same individuals.

Correlation coefficients range from $+1.00$ through $.00$ to -1.00. A coefficient of .00, of course, indicates the complete absence of relationship. The coefficients of 1.00 indicate perfect relationships, the signs merely designating direction. A correlation of $+1.00$ means that the relative magnitude of two traits corresponds throughout the full range of their distributions. An illustration of such a correlation is the relationship between the circumference of a circle and its diameter: the larger the circumference the greater the diameter, and vice versa. A negative correlation indicates an inverse relationship: for example, the erroneous concept of "a strong back and a weak mind." An illustration of a correlation of -1.00 is the relationship between the circumference of a wheel and the number of times it turns in a mile: the smaller the wheel the more times it turns, and vice versa. Seldom, if ever, do relationships of 1.00 occur in correlating physical, mental, and social traits, so it is much more commonplace to see such correlations as .87 between the strength of right and left grips, .42 between age and chest girth, and $-.59$ between standing broad jump distance and the amount of abdominal adipose tissue.

Product-Moment Correlation. The product-moment coefficient of correlation (r) is essentially a ratio related to the extent to which changes in two variables are associated throughout their distributions. "Moment" refers to the sum of the deviations from the mean (raised to some power) and divided by the number of cases. When corresponding deviations in two variables are multiplied together, summed, and divided by the number, the term product-moment is used. This form of correlation is most often used. It may be computed from grouped (scattergram) or ungrouped data, as long as paired relationships are retained. A large number of variations of the basic formula exist for correlating ungrouped data. One of these is particularly useful in making computations with electric or electronic computers.

The product-moment method of correlation assumes data which are continuous and present a rectilinear relationship. A rectilinear relationship exists when a straight line is the line of best fit for the paired scores throughout the distributions. A violation of this assumption reduces the amount of the correlation coefficient from its true value. When data are not linear, the curvilinear (eta) formula is used. If the linearity of the relationship is in doubt, both r and eta should be calculated and the linearity tested. If the distribution is found to be linear, r and eta are approximately equal. Eta is never less than r, but may equal or exceed it.

Reliability of r. In the discussion of reliability above, it was explained that a sampling distribution of means (as well as other statistics) takes on the form of a normal distribution. The whole concept of reliability was then based on the assumption of this normality. In a distribution of sampling r's, however, normality exists only if the true r is .00 and the sample size is large. With a high correlation, say of + .75, the distribution of sampling r's is negatively skewed and leptokurtic. As a consequence of this fact, a special problem is presented in applying tests of significance and in determining the probable limits of the true relationships.

In testing the significance of a correlation coefficient (whether or not the obtained r is due to errors in random sampling), the null hypothesis (that the true correlation is .00) may be applied at various levels of confidence. This may be done by formula or by use of a specially prepared table which provides the correlations which would reject the hypothesis at the 5 percent and 1 percent levels of confidence.

To determine the probable limits of the true correlation, the r should be converted into Fisher's z-function. The distribution of sampling z's is nearly normally distributed, so it is safe to apply a standard error of z in the same manner as for the mean. Once the z's for the limits selected are computed, the z's may be converted back to r's. The same process of z conversion is necessary when r's are averaged or when the significance of the difference between r's is tested. A table exists to make the r-z conversion easy (8:210).

Special Correlational Methods. Various methods of computing correlation have been devised for special conditions. The more common of these are described below.

Rank-Difference Method. The rank-difference method, designated as rho, is designed to determine the degree of correlation between two variables when ranked in order from high to low. For example, pupils may be ranked in order of merit for such qualities as sportsmanship, athletic ability, neatness of appearance, and the like, and the degree of relationship found between any two of the traits. To illustrate further, at one institution preparing physical education teachers, ten sports were ranked by the staff according to their over-all value for inter-collegiate competition and were ranked according to the prevalence of alumni coaching the sports; the resultant rho was .95.

The rank difference method may also be used when there are only a few scores by ranking the scores in order of merit. Product-moment correlation deals with the size of scores as well as position in the series. Rank differences consider only the position of the items in the series, making no allowance for the size of gaps between adjacent scores. Also, accuracy may be lost in translating scores into ranks, as gaps are created in the rankings when a number of scores, all the same size, receive the same rating. As a consequence, the artificial use of the method of changing scores to ranks is not recommended, unless used for exploratory purposes only.

The significance of rho may be determined by application of the null hypothesis in the same manner as for r.

Bi-Serial Correlation. Bi-serial correlation may be used when one variable is expressed as a dichotomy and the other is continuous. In using this method, however, it is assumed that the dichotomous variable is continuous and normally distributed and that the correlational relationship is linear. Examples of such dichotomies are athletic and nonathletic, socially adjusted and maladjusted, pass and fail a test, and the like. In these instances, the traits would be found in all probability to be continuous and normally distributed in a random sample, if it were possible to measure them in finer units. The bi-serial correlation is comparable to r, when the underlying assumptions are met. Its significance may be tested by the null hypothesis. The r to z conversion is not appropriate in this instance, as the standard error of the same bi-serial r varies in accordance with the proportion of continuous scores in each of the dichotomous classifications.

Tetrachoric Correlation. This correlational method may be used when both variables are dichotomies. The underlying assumptions and interpretations expressed for bi-serial r also apply to tetrachoric correlation. It may be said further that the standard error of tetrachoric r is one and a half to twice as great as for a corresponding product-moment r. As a consequence, the deliberate use of this method by artificially forming dichotomies from continuous data has the effect of discarding approximately half the data.

Phi Coefficient. The phi coefficient is designed for use when two variables are truly dichotomous and are concentrated as two separate points or into two distinct classes. Thus, the assumptions of continuous and normally distributed variables do not apply. Examples of such variables are the color of eyes as blue and brown, married and unmarried men and women, and in-school and out-of-school boys and girls. With the application of appropriate corrections, the phi coefficient may be used when both dichotomous variables are continuous or when one dichotomous variable is a true dichotomy and the other is continuous. Tests of significance may be applied through computation of chi square.

Contingency Coefficient. When two variables are grouped in two or more categories, the contingency coefficient may be used to determine the degree of relationship between them. This coefficient has special value in showing relationship when the categories exceed two but are less than the number that would appropriately be used in the product-moment method of correlation. Furthermore, no assumption of normality in the distributions of the variables needs to be made, unless an interpretation similar to that of r is to be made. This coefficient, however, is restricted in size depending upon the number of categories. When the number of categories is the same for each variable, a correction may be applied to bring this into perspective with r. The simplest test of significance for the contingency coefficient is by utilizaton of chi square.

Spearman-Brown Prophecy Formula. The Spearman-Brown Prophecy Formula is used frequently in the construction of tests on which several trials have been given. In securing reliability for these tests, the sum of scores on the odd-numbered trials may be correlated with the sum of scores on the even-numbered trials. However, as the length of a test affects its reliability, and the odd-even correlation represents only one-half of the entire test, a correction for the full-length test becomes desirable. The Prophecy Formula is designed to make this correction. It is also used to estimate test-retest reliability to be obtained from increasing the number of trials on a test item.

Partial and Multiple Correlation. So far, all measures of relationship that have been considered deal with two variables only. It is very useful many times to estimate relationship when three or more variables are concerned. Multiple relationships are frequently designated by their orders. The two-variable correlation is known as a zero-order correlation; a third variable in the correlation is a first order; and each additional variable increases the order correspondingly.

Partial Correlation. By means of partial correlation, a third (or more) factor may be held constant while determining the relationship between two factors which might be influenced by this third factor. The correlation of two variables contains common elements, in addition to the factors being related. The relationship, which is unique to the two variables, is found when the common influences are removed. The removal may be accomplished through experimental design by control of the common factors or through mathematical process by partial correlation. For example, Bovard, Cozens, and Hagman (2:375) report that weight of college men correlated .52 with shot-put distance; and also that height correlated .40 with shot-put performance and .58 with weight. The height influence, therefore, is found in the .52 correlation. With the effect of height partialled out, the correlation became .39. Thus, it may be seen that the correlation between shot-put distance and weight is materially reduced when the effect of height is held constant.

In computing partial correlations, it is assumed that the correlation is linear and that the influence of the partialled variable is the same for all components or levels of the variable. In this latter assumption, for example, does age, as a partialled variable, have the same effect at all ages? The reliabilities of the variables must be high; thus, the number of cases must be large, as chance fluctuations in the zero-order correlations have a marked effect on the resulting partial correlation. Care must be exercised in assuming causal relationships from partial correlations, as other factors still not understood may be the causative factor; i.e., the correlation is not always freed from common factors, as there may be other unknown common factors present.

Multiple Correlations. A multiple correlation coefficient (R) gives the correlation between a single variable, or criterion, and the combined effects of two or more variables. The R permits the selection of the most valid battery for forecasting a criterion. For example, McCloy used this method for determining the importance of age, height, and weight upon athletic performance. Franzen depended on this method to show the influence of social and economic factors on the health of the child. There are various ways of computing multiple correlations, but basically they depend on the zero-order and partial correlations. As a consequence, the underlying assumptions for these forms of correlation also apply to R.

Multiple correlation coefficients are never less than the highest zero-order correlation with the criterion, or dependent variable. The addition of experimental, or independent, variables may or may not increase R. The best situation for a high multiple correlation is found when experimental variables correlate well with the criterion and low with each other. When increases in R occur, the first increase from adding a variable to the correlation is the largest; the subsequent increases become smaller and smaller until the coefficient no longer increases with the adding of variables. This phenomenon of diminishing returns occurs only when the variables are included in their order of importance in the multiple situation.

The Doolittle solution of a multiple correlation has been frequently used by research workers. This method provides a means of solving multiple correlation problems with a minimum of statistical labor, especially when there are more than four variables involved. The Wherry-Doolittle method (9:435) is particularly useful when many variables are included in the correlational matrix. This method selects the tests of the battery analytically and adds them one at a time until a maximum R is obtained. In addition, there are a number of mathematical checks on the accuracy of the computations. Multiple regression equations may be computed as an extension of this process. It is possible, however, to compute by electronic machine a multiple correlation including all independent variables in the correlational matrix, and then test the significance of the beta coefficients in the multiple regression (10:339). Those variables that do not affect the coefficient of multiple correlation (i.e., are not statistically significant in the beta test) can then be eliminated.

Statistical significance of the multiple correlation coefficient not only depends on the number of scores but on the number of variables composing the correlation. The null hypothesis may be applied most simply by use of specially constructed tables which provide the R's necessary at the 5 percent and 1 percent levels of confidence for different sizes of samples and numbers of variables (9:426-28).

PREDICTION

Through regression equations, the individual's score or measure on one trait may be predicted when his score or measures are known on one or more other (independent) traits. It is also possible to analyze the relative contributions that each of the independent variables makes to the predicted variable, and to state in terms of proportions or percentages just what these contributions are.

Regression Equations. A multiple regression equation may be written for any number of variables; and may be written in deviation form, score form, and standard score form. The equation expresses the relationship between one independent or criterion variable (X_1) and any number of independent variables (X_2 X_3 . . . X_n).

When the equation is written in score form, the partial regression coefficients give the weight of scores for each of the independent variables. Thus, the effect of each of the independent variables in determining the criterion measure is indicated, although their relative importance is not revealed. When the multiple regression equation is written in standard score form, the partial coefficients are replaced by beta coefficients or beta weights. The beta coefficients yield additional information in that they indicate the *relative* importance of each of the independent variables in predicting the criterion.

The calculation of the standard error of estimate of the multiple regression equation gives a measure of the amount of error to be expected when the criterion measure is predicted from the regression equation. The standard error is large unless the multiple correlation upon which the regression equation is based is high. As a consequence, it is often misleading to calculate regression equations when the multiple correlation is low, especially if the equations are to be used to predict individual performances. With a low correlation, the standard error of estimate is so large as to make the predictions practically worthless.

The use of the multiple regression equation in score form is especially useful from a practical point of view, as the criterion can be predicted directly from the actual scores obtained in the testing. To illustrate this process, the following example is presented. With junior high school boys, a multiple correlation of .985 was obtained between the Rogers Strength Index, as criterion, and leg lift strength and arm strength (Rogers), as the independent variables. The multiple regression equation in score-form was

$$SI = 1.22 \text{ (leg lift)} + 1.23 \text{ (arm strength)} + 499$$

If a boy had a leg lift of 1200 pounds and an arm strength of 400 pounds, his predicted SI based on the equation would be 2455.

Predictive Index. Coefficients of correlation may be reduced to "percentages," or predictive indices. Actually, the predictive index is the reciprocal of Truman Kelly's coefficient of alienation. It is interpreted in terms of the percent of prediction value better than a "best guess." In physical education this index has been used to compare the relative predictive value of two or more coefficients of correlation. The technique is valuable to show the improvement in predictive accuracy to be expected when estimating an individual's standing on one measure from his standing on a second related measure. For example, an r of .40 would improve the prediction by 8.3 percent over a "best guess" and an r of .80 would improve the prediction by 40 percent over a "best guess."

ANALYSIS OF VARIANCE

The analysis of variance provides a technique for finding a valid estimate of the errors arising from the random selection of the subjects of several samples. Furthermore, it provides a basis for analyzing the effects of various "treatments" on the sample subjects.

The methods of the analysis of variance are based on the concept that, when several independent random samples are drawn from the same population, two independent estimates of the variance of the population may be found. The first of these estimates is derived from the variance of the sample means and the second from the mean of the sample variances. In the case where several samples have been drawn from the same population, the two estimates of the population variance would then differ from each other only by chance. Thus, a method is provided for testing the hypothesis that several independent random samples are from a common normal population.

The general procedures and major phases of an analysis of variance may probably be best explained through an illustrative problem. The simplest case is that in which three groups are involved. Consider the hypothetical situation in which 60 subjects have been divided at random into three groups of 20 each. The investigator is interested in determining the effect of three different kinds of motivation during a training period for improving arm strength. The three groups are assigned at random to the three experimental treatments (motivational techniques), and at the end of the experimental period, each subject is measured for arm strength.

At the beginning of the experimental period, the three groups were known to differ from each other in arm strength only by chance. Hence, the question now resolves itself as to whether the groups still differ from each other only by chance, or if the differences among them are too large to be attributed to chance. If the conditions of the investigation have been properly controlled, so that all subjects were treated alike except for the motivational techniques, and the differences are too large to have reasonably occurred by chance, there remains only the conclusion that there is a difference in the effectiveness of the motivational techniques.

The first step in the solution of the problem is to derive the two estimates of the population variance from the arm strength measures of the three samples, and then determine whether or not these two estimates differ significantly from each other. The appropriate test to be used is the Variance Ratio or F test. If F lacks significance, as evaluated by a predetermined level of confidence, the hypothesis that the samples are from the same population is regarded as tenable. This completes the analysis, and in the problem under consideration the conclusion is reached that the motivational techniques used are equally effective. If, on the other hand, F is found to be significant, it is indicative of real differences in arm strength among the groups, that is, differences too large to consider chance as the probable cause. In this case, the hypothesis that the samples are from a common population is rejected at the level of confidence established for the problem. It is now known that in terms of arm strength, the groups are no longer from the same population. This test does not, however, reveal between which groups the differences exist nor the exact cause of the significant F. The significant F may have resulted from differences among the means, the variances, or a combination of the two.

The primary concern of the investigator is to detect whether the groups differ from each other on the average, but, before a specific test for the means is applied, homogeneity of variance should be established. The best estimate of the population variance, to be used in the standard error of the difference between means formula, is the mean of the three sample variances. Therefore, it is necessary to demonstrate that the variances are homogeneous; otherwise, it would be incorrect to pool them. Also, if heterogeneity of variance is present, it may have resulted from correlation between the means and variances of the samples. Since the F test assumes independence of the two estimates of the population variance, this cause of heterogeneity of variance would render the F test invalid.

The second major step in the solution of the problem is, then, to apply χ^2 to test the hypothesis of homogeneity of variance. If χ^2

lacks significance, the conclusion is reached that the samples are from a population with a common variance and the significant F could have resulted only from real differences in means. If, however, χ^2 is significant, it would become necessary to transform the raw data to another scale in an attempt to stabilize the variances. If a suitable transformation was found, the entire analysis would then be performed on the transformed scores.

The final step in completing the analysis, after homogeneity of variance has been demonstrated, is to apply the t test for the purpose of locating the means between which real differences exist. The hypothesis being tested is that the samples, taken two at a time, are from the same population relative to arm strength. In cases where the hypothesis is tenable, it is concluded that the two motivational techniques compared are equally effective. In cases where the hypothesis is rejected, it is concluded that one motivational technique is more effective than the other, the relative size of the means indicating which is more effective. At least one significant t must emerge from this analysis and it is, of course, possible that all t's will be significant.

It should be clearly understood that valid interpretations from such an analysis depend on satisfying certain conditions: (a) The samples of the investigation have been drawn at random from a common normal population; (b) The variances of the treatment populations are homogeneous and the distribution of the treatment populations is normal.

The first assumption is satisfied by the random assignment of subjects to groups and groups to treatments. Homogeneity of variance may be tested by χ^2, but the normalcy of the population distributions may be difficult to demonstrate, owing to the small number of subjects usually available in experimental investigations.

This type of investigation is particularly applicable to methods experiments, but it may also be applied to a variety of situations. Typical examples of problems to be solved are the effect of psychological method on learning a motor skill, the effect of various distributions of practice or training in improving skill or physical condition, the influence of various diets in improving the nutrition of undernourished children, the effects of various bouts of exercise on fitness, the influence of recreational habits on work productivity.

The last example may be used to illustrate the application of the analysis of variance methods to observational data. The data collected in a study concerned with conditions which already exist in a population are referred to as observational data.

Let us suppose, for example, that an industrial concern wishes to explore the possibility that workers of different recreational habits also differ in their productivity. After recreational habits have been categorized, the population members (workers in the concern) would be identified according to their proper category. At this point there is a choice of procedures to follow. Each of the category groups may be treated as a subpopulation, and a random sample may be drawn from each of these subpopulations. The work productivity means of the individuals in the samples from the subpopulations are then tested by the methods of analysis of variance. An alternate procedure which may be followed is to select one large sample at random from the total population, divide the subjects into their proper categories, and apply the analysis to the means of the individuals in the various categories. A third possible procedure may merit particular consideration when the total number of subjects available is relatively small or the categories are so numerous as to make the subgroups quite small. In this procedure, all available subjects are divided into proper categories and the category means are analyzed.

One of the chief limitations of an analysis of observational data arises from the fact that frequently it is impossible to develop clear-cut interpretations from the results. This difficulty arises from the inability of the investigator to identify or control associated conditions as satisfactorily as in a controlled experiment.

ANALYSIS OF COVARIANCE

The analysis of covariance is an extension of the methods of analysis of variance in which the methods of regression and the methods of analysis of variance are synthesized.

The advantage of the analysis of covariance over the analysis of variance lies in the increased precision of the error estimate. It is quite possible that under the covariance analysis, because of the reduction of chance error, the conclusion from the analysis of variance will be reversed. The gain in precision is achieved through a statistical technique which equalizes the initial individual differences of the subjects with respect to the criterion trait, or some trait related to the criterion trait.

To achieve this equality, adjustments are made of the final criterion scores of the subjects on the basis of their initial scores. This adjustment of scores introduces a precision which is comparable to that obtained when groups are initially matched. There are, however, problems involved in matching subjects which are obviated by the methods of covariance. When groups are to be matched, the subjects cannot be assigned to the groups and time schedules cannot be planned until after the initial test is given. Under an analysis of covariance, subjects may be randomly assigned to groups immediately, since the modifications for initial status are performed after the treatments have been applied.

The effectiveness of the covariance analysis as compared to the variance analysis is dependent directly on the degree of relationship which exists between the criterion trait and the trait selected for initial measurement. If this relationship is small or negligible, no advantage will be gained and the considerable time and effort involved in securing the initial measure will have been wasted.

A brief explanation of the principles involved should at least serve to clarify the underlying theory.

The adjustment of the scores is made by determining the regression of the final criterion measures on the initial measures. Basic to this process is the assumption that there is one true regression of the final on the initial measures which is constant for all of the groups. If the deviation of an initial score from the mean of the initial scores and the regression coefficient are known, an estimate may be found of the value of the deviation of the final score from the mean of the final scores. Thus, it is possible to estimate for each subject how much his final score will deviate from the final mean, as a result of his initial status only, without regard for the treatment to which he was subjected. If this estimated deviation is subtracted from the subject's final score, it yields an adjusted score which expresses the individual's final status, with the effect of his initial status ruled out. If the mean of the adjusted scores is then found for each group, these means will not be affected by chance differences in the initial ability of the subjects with respect to the criterion measure or some measure closely related to it. These means will have the same relative size that they would have had in the case where subjects were matched at the beginning of the investigation.

Stated in the simplest way possible, it may be said that the analysis of covariance is, in effect, an analysis of variance of the adjusted scores. Actually, the statistical procedures are such that adjustments of individual scores are not necessary and the adjustments are made directly to the means.

The analysis of covariance could well be applied to the problem used for the purpose of illustrating the analysis of variance. Even though the subjects were initially assigned to the groups at random, the vagaries of chance are such that there may well be considerable variation in the arm strength of the subjects from group to group at the beginning of the training period. Since it is to be anticipated that chance variations in initial arm strength will be reflected in the final measure of arm strength, it behooves the investigator to take steps to equalize the initial arm strength of his groups.

There is no doubt that a substantial relationship will exist between the initial and final arm strength measures. This relationship is used in the covariance analysis to estimate the adjusted arm strength means, which as a result are free from chance differences in initial arm strength. Since the subjects have been made more alike by this procedure, there is a reduction in the variability of the adjusted scores as compared to the final scores. This decrease in the variability of the adjusted scores (residual variance) is reflected in the error estimate, which is correspondingly decreased.

It would be quite possible that, in the solution of the arm strength problem by the analysis of variance, the error estimate would be so large that the final conclusion reached would be that there is no evidence from the analysis that any of the motivational techniques is more effective than another. Owing to the reduction in the size of the error estimate, the same arm strength measures analyzed by the covariance methods could very well provide substantial evidence that there is a difference in the effectiveness of all of the motivational techniques, or between any pair or combination of pairs. The error estimates obtained from the two types of analysis are equally valid, but the estimate from the analysis of covariance has the added advantage of increased precision.

The reader may well ask why an analysis of variance would ever be performed when the precision of the investigation is so effectively increased by the simple expedient of securing an initial measure. The answer is that it is not always possible to identify a satisfactory initial measure. For example, if an investigator wanted to compare the effectiveness of three different methods of teaching swimming to beginners, he could not possibly obtain a measure of initial swimming ability, since the subjects have no swimming skills. He might rationalize several factors which conceivably could be related to success in learning to swim. However, in general, the contributing factors would be so numerous, or so complex, that it doubtless would be futile to anticipate that any one of them, or even two or three combined, would be sufficiently definitive to satisfy the demands of the analysis of covariance. In such a case the analysis of variance is indicated.

The general procedures applicable to the analysis of covariance are the same as described for the analysis of variance. The main difference is that the covariance analysis is in terms of the adjusted measures and demands certain additional calculations in order to develop estimates based on these adjusted measures. The development of the adjusted measures also introduces additional assumptions, and the validity of the interpretations, as in any analysis, will be directly affected by the degree to which the assumptions are met.

The first assumption is that the samples of the investigation have been drawn at random from a common normal population, and in addition all of the samples, in terms of the criterion measure, are random samples from their respective treatment populations. Both phases of this first assumption relative to randomness may be satisfied by the assignment of the subjects to the groups at random and of the groups to the treatments at random.

A second assumption is that the initial measures are not affected by the experimental treatments. If the initial measure is secured before the treatments are given, this assumption cannot be violated.

The third and fourth assumptions are concerned with the regression of the final scores on the initial scores. It is assumed that this regression is homogeneous for all of the treatment populations and that it is linear. Both of these assumptions may be tested—the homogeneity of regression by the F test and linearity of regression by χ^2.

The final two assumptions have to do with the distribution of the adjusted scores in the treatment populations. It is assumed that in each treatment population the scores are normally distributed and that the variances for the populations are homogeneous. The χ^2 test may be used in testing both of these assumptions. However, some difficulty may be encountered in testing for normality since in many experimental studies the number of subjects is too small to meet the conditions of such tests.

In addition to the simple analyses discussed, both the analysis of variance and covariance are basic to the solution of problems under more complicated experimental designs.

FACTOR ANALYSIS

Factor analysis provides a valuable aid in understanding the basic causal variables in any given field. If the primary or unitary abilities, designated as factors, are isolated by this technique, relationships will be revealed which may then be studied in greater detail by other methods.

Factor analysis offers a means of summarizing the complex interrelations which are present in a series of intercorrelations, allowing the determination of the smallest number of uncorrelated primary abilities which must be assumed in order to account for the table of intercorrelations. It provides a means of analyzing the factorial composition of each test, revealing the extent to which each independent ability is represented by each test. It permits the writing of regression equations which will predict the amount of any primary ability possessed by an individual, and it may be used to estimate the best combination of test items to predict the composite of all of the factors.

Test scores are converted into tables of correlations, and tables of correlations are converted into tables of factor loadings by statistical techniques which add nothing to the original data. The interpretation of the factors is not simply a statistical matter, but requires insight into the nature of the original tests.

Several methods have been devised for reaching the solution to a factor problem. Among the most prominent methods are those proposed by Spearman, Holzinger, Hotelling, Kelley, Tryon, and Thurstone.

Wolfle (21) suggests the following criteria for evaluating the different methods: accuracy of the original table of correlations, independence of factors, ease of computation, parsimony of the number of factors in the entire battery, parsimony of the number of factors in each test, goodness of geometric fit, consistency of factor loadings when a test is analyzed as part of a new battery, ease of interpretation of factors, and opportunities for testing one's hypothesis regarding a factor. On the basis of these criteria, Wolfle concludes that Thurstone's multiple factor analysis is the most satisfactory. Wolfle states his decision is in agreement with the opinion of Garrett, Guilford, Marginean, McCloy, and others.

Limitations of the factor analysis which are mentioned by Wolfle (21) are: (a) The solution is seldom a unique one; (b) The exact measurement of individual scores on the factors is possible only with the Hotelling technique; and (c) The factor pattern is dependent on the sample studied and may differ considerably from that found in the population.

A typical problem to which the technique of factor analysis is applicable is an analysis of the primary abilities essential to basketball players. The investigator concerned with this problem would attempt to identify all the physical attributes and motor and sports skills which presumably would be prerequisite to the game. He would then attempt to select simple, valid measures of each of these attributes and skills, and finally he would administer the battery of selected tests to a representative group of individuals.

The analysis of the data thus collected would involve developing a correlation matrix which includes the intercorrelations of all of the tests, development of the table of factors, and rotation of the factors.

From the table of rotated factors the factor identifications would be made. For example, in this basketball study, it may be anticipated that such factors as strength, speed, general co-ordination, hand-eye co-ordination, and possibly kinesthetic perception and peripheral vision would appear. The analysis would thus permit, among other things, the identification of the primary abilities basic to the battery of tests and, in addition, would indicate which primary abilities are most essential in the performance of each of the specific skills of basketball.

The method of factor analysis has already found numerous applications in the areas of health and physical education. It has been used, for example, in the analysis of growth, strength, anthropometric measures, cardiovascular measures, physical fitness, and a variety of motor and sports skills. In most instances, the Thurstone multiple solution has been used, with orthogonal rotations by the two-at-a-time method. A study by Cumbee (4), however, suggests that the multiple group method of rotation may be more advantageous than the two-at-a-time method.

NONPARAMETRIC METHODS

Many widely used statistical tests are limited in their application, since they assume certain conditions about the population from which the sample was drawn. These tests are validly applicable only in those cases where the assumed conditions are actually present in the parent population. This class of tests is known as parametric tests.

For example, one of the commonest assumptions is that the samples have been randomly drawn from a *normal* population. In situations where this condition is found, the appropriate parametric test should be used since it will be most efficient in the utilization of the data. In cases where the assumption of normality is not met, it is then desirable to use the proper nonparametric or distribution-free test.

Nonparametric tests are free from any assumptions concerning the distribution of the parent population (other than that it is continuous in certain tests), and the calculations involved in most of these tests are quite simple.

Several tests which have proven very useful in the past are of the nonparametric class. Such tests as the χ^2 test for goodness of fit, the coefficient of contingency, and the rank correlation methods are examples of this class of test. More recently, many additional nonparametric tests have been developed which are valuable in testing hypotheses about single samples and two or more samples, either related or unrelated.

Typical hypotheses which may be evaluated and some of the appropriate nonparametric tests are as follows:

1. The single sample
   (a) The population distribution from which the sample was drawn conforms to some theoretical distribution: the Kolmogorov-Smirnov one sample test (15:147).

2. Two independent samples
   (a) There is no difference in the distributions of the populations from which the samples were drawn: the Kolmogorov-Smirnov two sample test (20:426; 15:127); the run test (20:428).
   (b) The two samples were drawn from a common population: sum of ranks test (20:434).
   (c) The two samples were drawn from populations with the same median: the median test (20:435; 15:111).
3. More than two (r) independent samples
   (a) The r samples were drawn from populations with the same median: the median test (20:435; 15:179).
   (b) The r samples were drawn from a common population: analysis of variance by ranks (20:436; 15:184).
4. Two related samples
   (a) The median of the population of differences is zero: the sign test (20:430; 15:68).
   (b) The mean of the population of differences is zero: the signed-ranks test (20:432; 15:75).
5. More than two (r) related samples
   (a) The r samples were drawn from a common population: analysis of variance by ranks (20:438; 15:166).

MECHANICAL DEVICES

Many studies involve the collection of numerous data, the proper analysis of which demands extensive calculations. The time and effort required to organize the data and perform the calculations by hand may well be so exhorbitant that for the usual investigator the expenditure is prohibitive. Fortunately, there are mechanical aids available which so reduce manual and mental labor that hundreds of hours of time may be salvaged.

The automatic calculator greatly simplifies such tasks as adding, subtracting, multiplying, dividing, and extracting square roots which are sometimes ends in themselves and at other times basic to more extensive calculations. Through the use of the cumulative multiplying system, the sums of measures and the sums of squares of measures necessary for standard deviation may easily be secured. This system also greatly facilitates finding a zero order correlation. All of the essential terms, sums of measures and sums of squares for both variables, and sums of the cross-products are found simultaneously with no more effort than is involved in tabulating the paired values of the two variables. The tediousness of the calculations demanded in the solution of problems based on the analysis of variance is also greatly resolved by the use of a calculator.

In studies such as extensive checklist and questionnaire surveys, in the correction and analysis of large numbers of test papers, in multiple correlation, factor analysis, and analysis of variance, the use of an electronic computer affords immeasurable relief from the onus of handling and rehandling papers and performing massive calculations.

When it is known that an electronic computer is to be used, it is advisable for the investigator to consult with the individuals in charge of the computer. Considerable time and money may be saved if the data are recorded in the most efficient form for the purpose of transcription to the cards on which they will be punched. Companies have set up computing centers, with manuals of program abstracts which describe the various functions the computers will perform. If the investigator locates a program in which he is interested, he may send to the company for a copy of the complete program which describes in detail its purpose, method of functioning, limitations, precautions, etc. Copies of program details will also usually be available at the center where the computer is located.

Some of the simple functions which may be performed in organizing the results of a checklist or questionnaire study are sorting the cards by categories such as age, sex, school, grade, etc.; counting the number and type of responses to the various questions; providing information such as the number of persons who answered "no" to one question who also answered "yes" (or "no") to another; printing on sheets of paper any information on the card, in any desired arrangements; tabulating and printing sums for the total group and for any subgroups desired.

In a correlational analysis study with as few as 20 variables, 190 intercorrelations would be present, imposing a terrific calculational burden even if an automatic calculator is available. The electronic computer can provide the means and standard deviations for each of the variables, and the coefficients for all of the intercorrelations. In an analysis of variance problem, the sums of squares for the various error estimates are found by the computer.

BIBLIOGRAPHY

1. BLUM, J. R., and FATTU, N. A. "Nonparametric Methods." *Review of Educational Research* 24:467-87; 1954
2. BOVARD, JOHN F.; COZENS, FREDERICK W.; and HAGMAN, PATRICIA E. *Tests and Measurements in Physical Education.* Third edition. Philadelphia: W. B. Saunders Co., 1949.
3. COCHRAN, W. G., and COX, GERTRUDE M. *Experimental Designs.* New York: John Wiley & Sons, 1950.
4. CUMBEE, FRANCES Z. "A Factorial Analysis of Motor Co-ordination." *Research Quarterly* 25:412-28; December 1954.
5. EDWARDS, A. L. *Experimental Design in Psychological Research.* New York: Rinehart & Co., 1950.
6. EDWARDS, A. L. *Statistical Methods for the Behavorial Sciences.* New York: Rinehart & Co., 1954.
7. FISHER, R. A. *Statistical Methods and Scientific Inference.* New York: Hafner Publishing Co., 1956.
8. FISHER, R. A. *Statistical Methods for Research Workers.* No. 5. Eleventh edition. New York: Hafner Publishing Co., 1950.
9. GARRETT, H. E. *Statistics in Psychology and Education.* Fifth edition. New York: Longmans, Green & Co., 1958.
10. JOHNSON, P. O. *Statistical Methods in Research.* New York: Prentice-Hall, 1949.
11. LINDQUIST, E. F. *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin Co., 1953.
12. LINDQUIST, E. F. *A First Course in Statistics.* Boston: Houghton Mifflin Co., 1942.
13. LINDQUIST, E. F. *Statistical Analysis in Educational Research.* Boston: Houghton Mifflin Co., 1940.
14. MOOD, A. M. *Introduction to the Theory of Statistics.* New York: McGraw-Hill Book Co., 1950.
15. SIEGEL, SIDNEY. *Nonparametric Methods for the Behavioral Sciences.* New York: McGraw-Hill Book Co., 1956.
16. SNEDECOR, G. W. *Statistical Methods.* Fourth edition. Ames, Iowa: Iowa State College Press, 1946.
17. THURSTONE, L. L. "Current Issues in Factor Analysis." *Psychological Bulletin* 37:189-236; April 1940.
18. THURSTONE, L. L. *The Vectors of the Mind.* Chicago: University of Chicago Press, 1935.
19. WALKER, HELEN M. *Elementary Statistical Methods.* New York: Henry Holt & Co., 1943.
20. WALKER, HELEN M., and LEV, JOSEPH. *Statistical Inference.* New York: Henry Holt & Co., 1953.
21. WOLFLE, DAEL. "Factor Analysis to 1940." *Psychometric Monographs* 3:69; 1940.

CHAPTER 8

Construction of Tests

RAYMOND A. WEISS
M. GLADYS SCOTT

THIS CHAPTER MAY BE USEFUL IN AT LEAST FOUR WAYS. FIRST, it may serve as a guide to the person who wishes to construct a test. Knowing what qualities he seeks to measure, the test constructor must first devise and then validate a test that will evaluate these qualities.

Secondly, it may be useful for the person who wishes to verify or extend the validity of an existing test. Some fine research has been published wherein the authors have provided additional information about the reliability or validity of tests, or who have extended tests for use by groups not originally included when the tests were first developed.

This chapter may be useful, also, to the specialized person who serves as advisor to graduate students. Although test construction is a procedure requiring scientific skill, it is not out of reach of students who are supervised by qualified and experienced educators with good test construction backgrounds.

Finally, this material may be useful to the teacher who wants to know how tests are constructed so that he can evaluate the tests he uses in his own program. Knowledge of test construction enables him to read reports about tests and to judge the research procedures used by the authors of the tests.

REASONS FOR CONSTRUCTING TESTS

Compared with all those who use standard tests, only a very small number of people ever develop the tests they use. The test constructers and the test users are two distinct groups; and, as one might guess, users of tests often have other reasons or different conditions in mind than did the authors at the time the tests were constructed. Consequently, many tests are suitable for use only under special or unusual conditions. The following examples demonstrate this idea.

For Research. Often a research worker finds that the instrument he needs for his research project does not exist; therefore, he has no alternative but to develop his own device. His major objective is to produce an instrument that is accurate to a high degree of refinement. He cares less about practical considerations, such as the amount of equipment needed, expense of equipment, the number that can be tested in a unit of time, and space requirements. Consequently, some tests that are useful in research, such as an endurance run on a motor-driven treadmill, would be impractical in a school testing program.

For a Local Situation. Sometimes, a teacher wants a tailor-made test to fit his own situation and is willing to spend the time and effort needed to develop such an instrument. The result may be just what the author wants but may not be nearly so suitable in other programs. To illustrate, the author of a knowledge test may plan the content of the test to match the content of the health education course he teaches. This is natural enough, but if the author plans to publish his instrument, other users would have to be sure they agree with that content.

For Special Groups. Tests have been constructed for such special groups as the military services, recreational organizations, boys' clubs, social agencies, and state educational programs. In most cases, each test is adapted to the special purposes and characteristics of the organization. To illustrate, the Air Force physical fitness test was planned so that it could be administered to large numbers in short periods of time. Recently, a physical performance test was adopted for use in one of the states. As part of the development of this test, items were selected only after they were judged adaptable to the special conditions of the school program in that state.

Such adaptation to special groups is commendable, but it is obvious that another group interested in such a test would first have to determine its suitability to that group's own program.

Faced with all these special circumstances it is little wonder that authors of tests have developed instruments that seem more suitable for one situation than for another. In some ways, this is unfortunate. It is probably true that teachers who do not conduct testing programs are influenced more by the unsuitability of available tests than by their lack of validity or reliability.

Because the materials in this section may have some bearing upon the construction of future tests, the need to develop tests that are administratively feasible and widely adaptable cannot be overstressed.

TYPES OF TESTS

Although there are many more than two types of tests, it is convenient to place them all into two large categories when discussing test construction procedures—namely, written tests and physical performance tests. Whereas physical education and recreation use both written and physical performance tests, health education, on the other hand, has more use for the written tests. In fact, tests of proficiency in first aid are probably the only health education physical performance tests, and these bear little resemblance to motor tests in the other two areas.

The remaining materials in this chapter are presented in two sections—one dealing with written tests and the other, with physical performance tests.

Written Tests

Tests that measure *knowledge* and *understanding* are used in all three areas of health education, physical education, and recreation. The reader should know that knowledge is not the same as understanding. Knowledge requires simple recall and is based on memory. On the other hand, a person displays understanding when he shows that he can use his knowledge to form reasoned judgments. A student shows knowledge when he responds, "nine players," to the question, "How many men on one team take the field in a standard baseball game?" However, understanding is required to answer the question: "What is the meaning of a news report on a tennis match between A and B which reads (7-5), (6-4), (6-8), (8-6)?"

(a) A close match between women
(b) A close match between men
(c) An easily won match between women
(d) A deuce match between men
(e) A close match, but cannot tell whether it is men's or women's match.

To answer this question correctly, the student would have to recognize the relationship between such bits of knowledge as game scores in a set, difference in number of sets for men and women, meaning of sequence of scores, number of sets required to win, and the usual form of match reports.

Another form of written examination is the test of misconceptions. These tests can be developed in any of the three fields, although health education has developed the most, apparently as a means of ferreting out and eliminating health misconceptions. The question form usually consists of true-false statements, with the significant ones being false. True statements are scattered throughout the examination to camouflage the purpose of the test. Sample statements are: "An overweight person can be confident he is not malnourished." "Insanity is a shameful disease which is a sign of evil." "Accidents usually happen to unlucky people."

Still another form of written test is the instrument for measuring attitude. Here, the person who responds to each question shows his feeling about the importance or value of the topic contained in the question. For example, the student who responds, "Agree," to the statement, "I would take physical education only if it were required," reveals that he does not place much value on physical education as a subject in the school.

Although the explanations above cover most types of written questions, sometimes authors use different terminology in naming the tests. For example, some are labeled tests of status, others are called tests of habits, still others are tests of interests, and so on.

Although different in content, these tests are all constructed along the same general procedures: (a) preparing the preliminary form of the test; (b) testing for validity; (c) preparing the final form of the test; (d) testing for reliability; and (e) preparing norms.

PREPARING THE PRELIMINARY FORM

The purpose of this step is to produce a preliminary form of the test which can be experimented with in the succeeding steps. Although the first draft of any test seldom remains unchanged, the experimenter tries for perfection from the start. He knows that he can save time in the later steps by taking extra care in developing the preliminary form. The experimenter develops the preliminary form of the test in at least five steps: constructing the framework, collecting statements of information, choosing the form of the questions, preparing the questions, and preparing the instructions and other information needed for the examination form.

Constructing the Framework. A written test can cover as much or as little content in a subject as the author wishes, and he must decide upon the scope of the examination at the start. This scope usually depends upon the purpose of the test. A test that measures achievement at the end of a course will cover a greater scope than one which tests for learning in one unit of a subject.

The test author identifies the scope of the test by listing all aspects of the subject he wishes the test to cover. This listing is called the test framework, or the table of specifications.

If the author is preparing a test to use in his own class, he may apply his course outline as the framework for the test. This is logical, as he would naturally want the test to cover the same elements he covers in his course.

However, if the experimenter is developing a standardized test, he will want to search authoritative sources, such as books and articles, to be sure his framework covers all important aspects of the subject.

The framework should also show the relative importance of each item in the listing. This may be used as a guide in determining the number of questions for each item of the framework.

The procedure used by Langston (25) in standardizing a volleyball knowledge test demonstrates the development of a test framework. In compiling a list of the most commonly emphasized phases of volleyball, he reviewed 40 articles written about volleyball, 14 of the latest issues of the *International Volleyball Review,* and the four latest issues of the *Official Volleyball Guide.* From an analysis of these materials, Langston identified 11 distinct parts in the game of volleyball. Next, he asked a jury of experts to rate the importance of these 11 phases. In this way, he achieved the following framework of his volleyball test:

Volleyball Phase	*Percentage Value*
History	3
Pass	11
Set-up	11
Spike	12
Net recovery	8
Block	9
Service	10
Offensive strategy	12
Defensive strategy	11
Rules	8
Officiating	5

Kelly and Brown's study (23) to develop a test of field hockey provides another example of the weighted framework. In surveying 11 textbooks, they found that coaches or teachers of field hockey should have competence in four areas—rules, techniques, coaching procedures, and umpiring. Within these four areas, the literature revealed 82 topics which Kelly and Brown weighted from 1 to 5 in the order of increasing importance and complexity.

Collecting Statements of Information. The content of any test is developed from the subject matter of the study unit or units to be covered in the test. The test author prepares statements of information to cover all phases of subject matter listed in the test framework. He may collect these statements from his lecture notes, from textbooks, from library reference materials, and even from his own knowledge and understanding which he has gained during his professional career.

Two factors influence the number of statements the author prepares. One factor is an anticipated loss of some statements which will be discarded later if they are not found to be valid for the test. There is no set rule for estimating this kind of loss, although it is not uncommon to prepare double the number of questions needed on the final form. For example, Langston (25) decided to start with 206 questions after fixing the final size of the test at 100 questions. This last item, the final size of the test, is the second factor. Because the number of statements determines the length of the examination, the author must now decide upon the time limit for the test and the total number of questions a class can answer in that time. This total becomes the size of the final test, and the experimenter must work backward from that number in deciding how many statements to prepare at the outset. Phillips (36) found that college women could answer 100 items in a 50-minute period.

As he compiles the statements of information, the author groups them under each part of the framework. When he finishes collecting, he then screens the statements within each part of the framework, eliminating the less desirable ones and eliminating overlap or duplication.

Choosing the Form of the Questions. Before the test author can prepare test questions from the statements of information described above, he must first decide upon the form of the questions. Although there are many ways to plan questions, the four question forms used most are multiple choice, true-false, recall and matching.

Multiple Choice. The multiple choice question consists of a statement followed by several alternate responses (usually from 3 to 5), one of which is the correct or best response. Kilander (24) uses the following type in his *Health Knowledge Test for College Students:*

Which disease is transmitted most readily and quickly by personal contact?
1. Cancer
2. Pellagra
3. Nephritis
4. Anemia
5. Diphtheria.

The *Health Practice Inventory*, by Johns and Juhnke (18), asks how often the individual uses various health practices. For example:

Do you participate in walking or hiking as part of your daily activity program?
1. Never
2. Rarely
3. Sometimes
4. Usually
5. Always.

The test author can use any type of multiple choice response he wishes, depending upon the content of the question. The above are two of many kinds.

True-False. In answering true-false questions, the student agrees or disagrees with the statements. Here is a common form for true-false questions:

T F 1. A person increases his strength by increasing the number of muscle fibers through exercise.
T F 2. In soccer a goal may be scored directly from the kick-off.

The student marks each question true or false by circling the T or F preceding the question.

Recall. In answering recall questions, the student writes the answer in a blank space provided on the examination sheet. The answer may be one word, a phrase, or even a sentence. Here are examples of two types of recall questions.

1. How many bones are there in the body? ________ ——
2. List four tests of physical fitness: (a) ——
(b) ——
(c) ——
(d) ——

Matching. In answering matching questions, the student compares two columns of words and matches each word in one column with the appropriate word in the second column. Often, one column contains phrases rather than single words. Here is a partial set of matching statements.

Select the best statistical procedure for each occasion:

—— When interested in the dispersion of scores.	1. Standard scores
—— When comparing one group with another.	2. Correlation
—— When looking for a substitute of one test item for another.	3. Standard deviation
—— When interested in computing a composite score for a multi-item test.	4. t test
	5. Histogram
	6. Step interval

Which To Use? Test construction experts agree that the multiple choice type is the most valuable of the objective question forms. It is especially good when testing the students' understanding and ability to judge and discriminate. Tests in physical education and health education employ mostly multiple choice questions.

The recall type of questions is rated as valid, but tends to encourage rote learning. In preparing for examinations containing recall questions, the student is encouraged to memorize isolated bits of information. This type of question cannot be scored by machine.

Matching questions, like recall questions, tend to encourage rote learning. Also, they are more difficult to construct than other types of questions. Unless skillfully prepared, this test form is likely to provide clues to some of the correct responses.

The true-false type is the least reliable of all the objective forms, and educators discourage its use except where other forms do not fit the situation. Some educators object to this form on the basis that it encourages guessing, and others object to questions which are false, saying that they have an effect of negative suggestion.

There is a definite tendency now to favor multiple choice questions for objective written tests. However, the other forms may be useful in some instances, provided that the questions are carefully prepared.

Attitude Test Questions. The statements in attitude tests are similar to the statements in true-false and multiple choice questions. However, the form of the response to the attitude statement depends upon the technique used in scaling the instrument. Two scaling methods prevail—one by Thurstone and the other by Likert. In the method developed by Thurstone (44), the subject is asked to select only those statements with which he agrees. Carr (6), in her attitude test, uses the Thurstone method by asking the subject to agree or disagree with each statement, as in the following examples:

I like quiet games with no running or jumping.
1. Agree
2. Disagree

In the method proposed by Likert (27), the subject selects one of several possible responses to each statement, as in the following example:

I enjoy a shower at the end of the physical education period.
1. Strongly agree
2. Agree
3. Undecided
4. Disagree
5. Strongly disagree

In constructing an attitude scale, the Thurstone method requires more time and labor than the Likert technique, but studies show that the two types of scales give substantially the same results. In recent years, authors of physical education attitude tests have preferred the Likert scaling method. Wear (48), Kappes (21), and McCue (28) developed their tests of attitudes in physical education in the manner of Likert.

The reader is referred to Chapter 5 for more information about attitude scales.

Preparing the Questions. After the test author has decided upon the form of the questions for the examination, his next step is to prepare questions to cover the statements of information he had selected earlier. It takes skill to write questions that are direct, clearly and concisely worded, and adapted to the difficulty level of the group for which the test is constructed. Scott and French (41: Chapter 8) list the following rules for constructing the four forms of questions, which are reproduced here with the permission of the authors.

Multiple Choice

1. Use a short, simple, direct question form for the stem.
2. Avoid choices which are not plausible or which are too obvious.
3. Avoid having more than one correct response if the directions call for selecting the *one* correct answer.
4. Avoid answering one question with another.
5. Avoid unintentional clues, such as placement of correct or best response consistently in a certain place in the series; word matching between the stem and the response; making the correct response consistently longer or shorter than the incorrect ones; the grammatical clue of using a singular expression in the stem and plural ones in all but the correct responses; and the grammatical clue of using an incomplete statement in the stem, ending in "a" or "an."
6. Avoid use of textbook language and of stereotyped phrases if the purpose is to test for ability to use information and for understanding rather than memorization. Use familiar or stereotyped phrasing in an incorrect response occasionally to deliberately mislead the shallow thinker.

True-False

1. Make the statements or questions brief and direct.
2. Avoid ambiguities.
3. Avoid textbook wording.
4. Have an approximately equal number of each alternative, with no regular pattern to responses.

Recall

1. Be sure only one word or phrase can answer it correctly.
2. State questions to elicit the briefest answers possible, for objectivity.
3. Provide spaces of uniform length, long enough for the longest reply, to avoid unintentional clues.
4. Provide space for answers in or near a margin to facilitate scoring.

Matching

1. The right-hand column should contain the responses and should always have at least two more items than the left-hand column, to prevent answering the difficult ones on the basis of elimination alone. The items in the right-hand column should be numbered or lettered.
2. Place blank spaces for recording the number or letter of the matching item in front of the items in the left-hand column.
3. Avoid clues in grammatical form or the use of proper names or capitalization.

4. Make the content of each list homogeneous.
5. State in the directions whether items in the right-hand column may be used more than once.
6. Be specific in the directions as to the basis upon which connections are to be made.
7. Arrange the left-hand column in sequence, alphabetically or numerically.

Preparing the Test Instructions. The test author should be as careful in preparing the test instructions as he is in preparing the test questions. These instructions should appear on the experimental test form so that the investigator can check their accuracy and clarity during the development of the test.

Space is usually provided at the top of the test form for the name, date, instructor, and name of the course. Some test forms call for other information such as age, grade level, name of school, date of birth, sex, father's occupation, religious affiliation, country of birth, and place of parents' birth. Investigators use this additional information in preparing norms and in studying factors that are related to achievement.

The directions should explain the mechanics of answering the test questions. When the test has more than one type of question, separate directions for each part may be needed. The time limit should be stated—or the fact of no time limit, if this is the case. Instructions often offer helpful hints such as "Do not spend too much time on any one question," and "Do not guess." Broer and Miller (5) use the following directions for parts of their tennis knowledge test:

> Part I. *Multiple true-false:* If the statement is entirely correct, encircle the "T." If the statement is totally or partially incorrect, encircle the "F."
>
> Part II. *Multiple choice:* Place a figure X opposite the statement which best applies to the particular situation.
>
> Part VI. *Matching.* The descriptions in Column II apply to some of the words or phrases in Column I. Place the appropriate letter from Column I in the blanks provided in Column II.

Sometimes, the test directions are illustrated by sample questions. The correct answer to the sample is marked according to the instructions.

Curricular Validity of the Test. After carrying out the preceding steps, the investigator now has a preliminary form of his test. Although he has no evidence of the statistical validity of the test, he should be able to demonstrate that the test does have curricular validity.

Validity is the degree to which the test fulfills its purpose. For example, an informational body mechanics test is valid if it really measures the students' knowledge and understanding about the efficient use of the body. Validity can be demonstrated subjectively and objectively; the former is termed *curricular validity.*

A simple way to demonstrate curricular validity is to describe in detail the process of constructing the test. The author can offer proof of curricular validity at each of the several test construction steps. In the first step (setting up a framework), the test will be valid to the extent that the items in the framework cover the subject of the test. If the literature shows that throwing, catching, running, and batting are important aspects of baseball, then the investigator demonstrates curricular validity when he lists these items in the framework and uses the literature to document their importance.

If the test author intends to use the test in his own course, he can demonstrate curricular validity by showing that the items of the framework cover the statements of the course objectives and the course outline.

The investigator can extend the evidence of curricular validity by showing that the statements of information used in preparing the test questions amply cover the content of the course or subject. In this connection, the author should show that the proportion of questions for each part of the framework places a proper ratio of emphasis upon that phase of the subject. For example, if the subject of softball should be taught with the following proportions of emphasis,

Rules	25%
Techniques	25%
Game strategy	30%
Team organization and tactics	15%
Safety, use of equipment, etc.	5%

then the statements on the examination should cover the subject matter in these proportions.

Still another way to demonstrate curricular validity is to show that each test question was constructed with care and in accordance with accepted criteria on rules such as those listed on pages 222-223. After the questions have been prepared, these same rules may be used to judge how well the questions turned out.

Although the author of the test may carry through the steps of curricular validity without assistance from others, in addition he may submit the preliminary form of the test to experts for verification. In this case he asks a committee of judges to evaluate the content of the test by considering such factors as:

1. Due consideration for functional values rather than exclusively factual content material
2. Importance of the respective items
3. Clarity and apparent efficacy of each item
4. Suitability of the item level of difficulty for the group to which it will be administered
5. Similarity of test items to situations in which abilities will ultimately be used
6. Proper ratios of emphasis as opposed to over-emphasis of certain points.

Also, the judges may be asked to evaluate the mechanical aspects of the test, such as:

1. The items should be direct questions, clearly and concisely worded.
2. There should be no clues or artificial sources of aid to any student.
3. There must be a real basis for the correct response.
4. The form of the item should be suited to its content and function.
5. The directions should be simple and understandable.
6. Possibility of errors in answering and scoring should be minimized by provisions for response.
7. The items should be well arranged from a psychological viewpoint.

The test author may be able to make valuable changes in the examination upon the suggestions of his committee of experts.

Finally, it may be said that curricular validity is demonstrated when the proper emphasis is assigned through appropriate proportions of questions for each part, and when the items in the test are carefully constructed.

To gain the details of test construction, which limitations of space make impossible here, the reader wil find thorough treatment of the subject in such references as Adkins (1), Ebel (9), Engelhart (10), Mosier, Myers, and Price (33), Ross (39), Scott and French (41), and Weitzman and McNamara (50).

TESTING FOR VALIDITY

Once the preliminary form of the written test is completed and its curricular validity demonstrated, the next step is to administer the test to a group of subjects and then analyze the scores item by item for statistical validity. Because evidence of test validity is affected by the type of subject used in the experiment, the test constructor should use subjects from the same population for whom the test is intended, and also select subjects with a range of ability in the test subject from high to low. Statistical validity of written items is usually tested in three ways: (a) index of discrimination, (b) difficulty rating, and (c) functioning of responses.

Index of Discrimination. For a test to be useful, it should distinguish between those who possess knowledge and understanding and those who do not. This is test validity. However, the ability of a test to distinguish between students of varying abilities depends upon the discriminating power of each item in the test. A test item is said to discriminate when the students who answer it correctly are found to achieve higher on the total test than the students who answer the item incorrectly. Several techniques are available for testing discriminating ability of test items. Some of these procedures are described below.

Flanagan Index of Discrimination. The Flanagan technique (12) yields a product-moment coefficient of correlation which indicates how well a test item differentiates good and poor performance. The correlation coefficient is high when the item is answered correctly by those who score high on the total test and answered incorrectly by those who score low on the total test. When high and low scorers do equally well on a test item, the item coefficient is low. Depending upon how the high and low scorers achieve on a test item, its index of discrimination may fall anywhere between high and low. In the Flanagan technique, a validity coefficient is computed for each item in the test. Thus, 50 validity coefficients are needed to validate a 50-item test.

In addition to the original reference describing this technique, Scott and French (41:286-92) describe the steps in detail for computing the Flanagan index of discrimination and introduce a sample worksheet that facilitates the computations involved. The experimental data needed are the scored examination papers, written by a group of subjects, showing the right or wrong mark for each question and the total score on each examination paper. The experimenter discards those papers in the middle of the total score distribution and works with the high and low score papers, usually the upper 27 percent and the lower 27 percent of the total number writing the examination. [See Kelley (22).] In analyzing a test item, the investigator computes the percentages of the lower and upper groups that answer the item correctly. For example, if 3 out of 30 low scorers answered an item correctly, the proportion of success would be 10 percent. Likewise, if 21 out of 30 high scorers answered that same item correctly, their success would be 70 percent.

The investigator then enters these two percentages in a special table, prepared by Flanagan, and reads off the index of discrimination. This index is an estimate of a correlation coefficient between success on the item and success on the criterion of total score. In the example given above, the index of discrimination turns out to be .6[illegible].

In order for an item to be retained in the same form, it should yield an index of approximately .20 or better. Although .20 is the minimum acceptable level, each investigator will have to decide upon the exact cutoff point, depending upon the subject matter involved.

The Flanagan technique is a simple procedure which has the advantage of

1. Dealing only with two divergent groups
2. Making use of a conversion table
3. Making it unnecessary to use each person's total score in the computation of the index.

Studies by Dzenowagis and Irwin (8), Kelly and Brown (23), and Hennis (17) illustrate the use of the Flanagan technique for testing item discrimination.

The Aschenbrenner Technique. Aschenbrenner (3) suggests that a smaller percentage of high and low scores be used in computing an index of discrimination when very large numbers of subjects write the examination. Instead of selecting the upper and lower 27 percent of the papers in the distribution, he suggests using the extreme 10 percent in the distribution tails. The index is computed in a manner similar to the Flanagan technique from a special table for the 10 percent extremes. Aschenbrenner suggests that the index be computed on no less than 100 papers and no more than 600. Hennis (17) used this technique in validating knowledge tests in physical education.

The Davis Technique. Davis (7) presents a technique in which the index of discrimination for a test item is expressed on a scale ranging from 0 to 100. He converts the Flanagan r into Fisher's z which, in turn, he converts into an index ranging from 0 to 100. This technique is more complex than the Flanagan technique, but except for this drawback it appears to have the same advantages. There appear to be no tests in our special fields which were validated using the Davis technique.

The Swineford Procedures. Swineford (43) presents two procedures for judging the validity of test items.

1. Swineford uses the test papers that fall into the upper and lower quartiles of a distribution and applies the following formula in evaluating each test item:

$$\frac{(R_U + W_L) - (W_U + R_L)}{\frac{N}{2}}$$

where R = number right, W = number wrong, U = upper quartile, L = lower quartile, and N = total number of papers in the distribution. The larger the index, the greater the discriminating power of the item. Perfect discrimination would yield an index of 1.0. Items should be rejected when the index falls below .4 or .5. A report by Scott (40) illustrates the use of this procedure.

2. In a second technique, Swineford evaluates item validity using the entire distribution of examination papers. First, she sorts the set of papers into two piles: one in which an item is answered correctly, and the other in which the item is answered incorrectly. Then she averages the total scores in each set of papers. Where these two averages differ by at least an amount equal to one standard deviation of the total group of papers, the item may be considered valid. The investigator repeats this procedure for each test item.

Scott and French (41:285) present a worksheet on which is tabulated a frequency distribution of correct responses to an item and another distribution of incorrect responses to the same item. The averages are then computed using these frequency distributions.

Although the Swineford procedure is more time consuming than the Flanagan and Davis techniques, nevertheless it is satisfactory in the absence of the latter's charts. The Swineford technique is illustrated in reports by French (15) and Scott (40).

The Votaw Technique. Votaw (45) also uses the upper and lower tails of a distribution to evaluate the validity of test items. However, unlike Flanagan and Davis whose index of discrimination is expressed as a correlation coefficient, Votaw compares the percent of high scorers who answer an item correctly with the percent of low scorers who also answer the item correctly, and then tests the significance of this difference using a probable error term.

Votaw also uses a formula which tests the validity of an item in a slightly different, but equivalent, way. With this formula, the investigator can estimate the proportion of the high scorers who would have to answer an item correctly when a known proportion of the low scorers answer the item correctly, if the item is to be judged valid. The investigator then compares this estimate with the actual proportion of high scorers answering correctly. If the actual proportion exceeds the estimate, the item is judged to be valid. Votaw suggests that this second formula be used to prepare a graph for interpolating the validity of test items. Since only 10 to 18 formula computations are needed to establish the curve, anyone planning to test the validity of many more than that number of items would save t. .e by preparing and using the graph instead of working the formula for each of the test items. Reports by Miller (31) and Phillips (36) illustrate the use of the Votaw technique.

The Phi Coefficient. Jurgensen (19) uses the following formula to determine the validity of a test item:

$$\phi = \frac{p_u - p_l}{(p_u + p_l)\quad(2 - p_u - p_l)}$$

where ϕ = phi, p_u = proportion of N in upper group which has correct answer, and p_l = portion of N in lower group which has correct answer. Phi is a value that can be converted into a critical ratio or a chi square. It also can be transformed into a value corresponding to a Pearson r.

Jurgensen has prepared a table to enter with proportion values, making it possible to find phi without using the above formula.

Once values of phi are computed, they can be used again and again regardless of the number of cases in the high and low scoring groups. This is an advantage over the Votaw formula where the results hold only for the N used in the formula. Broer and Miller (5) and Fox (13) used the phi coefficient in validating knowledge tests.

Validating Against External Criteria. All the indexes of discrimination discussed above have one thing in common. They are used to validate each item in a test against the criterion of the total test; hence the term, item validity. Accordingly, the validity of the test depends upon the validity of each test item. Thus, the validation procedure never breaks out of the circle, and it is clear that we are dealing with a form of internal consistency. An alternate to the use of such an internal criterion would be to validate the proposed test against some external criterion.

Although external criteria are frequently used in developing physical performance tests, they are seldom used with written tests. In developing a knowledge test for tennis, Broer and Miller (5) showed that intermediate tennis players scored higher on the test than did beginning players. The authors state, "A comparison of the two distributions can be used as an indication of validity of the total test." Here the criterion is external because the students were classified into two groups on the basis of a factor (experience) other than the test results. Phillips (36) also demonstrated that divergent groups scored differently on a badminton knowledge test. She found that major students in physical education scored higher than nonmajors who were classified as beginning and intermediate students.

Besides the divergent-groups criterion, the test constructor may use other types of external criteria. He may ask judges to rate students for their knowledge in observable class situations and then correlate the scores on the experimental test against these ratings. Kelly and Brown (23) correlated scores on a field hockey test for physical education majors with the instructor's rating of the students' competence to teach field hockey. Because judges' ratings are subjective, this technique must be used with great care. Kelly and Brown illustrated another kind of external validity criterion when they correlated test scores with the extent of the students' field hockey experiences.

In all of the above examples of external criteria, the authors were providing added evidence of test validity; their major source of validation was the item validity analysis.

Item analysis is a more appropriate form of written test validation than is the use of external criteria. Anyone planning to develop a written test would do well to use one of the several discrimination indexes available in the literature rather than some external criterion, although the experimenter may use an external validity procedure as an additional step.

Difficulty Rating. A test question is difficult if most students fail it, and easy if most respond correctly. The percentage answering correctly is the *difficulty rating* of the item. Thus, a question which is answered correctly by 40 students out of a group of 100 will have a difficulty rating of 40. This question would be rated as more difficult than a question with a difficulty rating of 90.

Items which are too difficult or too easy are likely to have low discriminating power. The experimenter should set limits and then discard items which exceed those limits. For example, those items with less than 10 percent or more than 90 percent difficulty ratings are frequently dropped. Because of the effect of difficulty upon validity, there appears to be some advantage in having the difficulty ratings concentrate around 50 percent. On the other hand, a spread between 10 and 90 percent will tend to insure discrimination at all levels of ability. Easy items help to discriminate among the poorer students, while difficult items prevent the top students from clustering on the grade scale.

The experimenter may not find it easy to get the range of difficulty ratings he seeks. In any event, he should be aware of the general level of difficulty of the test, as this information will help him to set standards for grading the students.

In developing their tests, Hennis (17), Langston (25), and Waglow and Rehling (46) illustrate use of the difficulty rating of test items.

Functioning of Responses. In multiple-response questions, although only one answer is "correct," the other responses to the question should be plausible enough to distract the student who does not really know the answer. When all the responses of a question are equally plausible, the student who does not know the answer is faced with the choice of omitting the answer or guessing. On the other hand, if the student knows that one or more of the responses are not correct (because they are implausible) he may be led in the direction of the right answer by having fewer plausible answers from which to choose. All the distracting responses to a question should function, and the test constructor should demonstrate this attribute as part of the validity of the test.

Testing for nonfunctioning distractors is a simple procedure. The investigator computes the percentage of students who select each response to a question. For example, if 60 students out of a group of 100 select the four wrong answers to a five choice question as follows,

1st distractor—8 students
2nd distractor—20 students
3rd distractor—25 students
4th distractor—7 students

the percent of responses would be 8, 20, 25, and 7 percent, respectively. Scott and French (41:282) suggest that any response not selected by at least 3 percent of the total number of persons taking the test be discarded as a nonfunctioning distractor.

Both Hennis (17) and Langston (25) used the standard of 3 percent in testing for nonfunctioning distractors. Hennis discarded any item with less than two functioning distractors. Langston either replaced nonfunctioning responses or dropped the entire question. Kelley and Brown (23) judged a response to be nonfunctioning if it was not selected by at least 2 percent of the subjects.

PREPARING THE FINAL FORM

In the preceding section, we saw how the test constructor tests the validity of the test items, using an index of discrimination, a difficulty rating, and a measure of the functioning of responses. Now, the experimenter is ready to reduce the original, experimental form of the test to its final form. He does this mostly by eliminating certain of the test items. If the author can discard all the undesirable items and still have a valid, well-balanced test, he will not need to revise any items, avoiding a repetition of the experimental testing.

The most important basis for discarding questions is item validity. The author tentatively sets aside any items that fail to meet the standards for discrimination, difficulty, and functioning of responses. Then he makes adjustments in the remaining group of questions so as to retain curricular validity, keep a balance among the types of questions, and end up with the desired length of test.

In retaining curricular validity, the author need only to be sure that the final form of the test has questions, in the correct proportion, to cover every part of the original table of specifications.

If the test contains two or more types of questions, the author may want an equal number, or some other proportion, of each type. For example, the author may want half the questions to be true-false and the other half, multiple choice. Here, the test constructor has to be careful in discarding questions not only to give the correct balance in types of questions but also not to disturb the proportion of questions needed for curricular validity.

In discarding questions, the author keeps in mind that he wants the final form of the test to be short enough to be administered within a given period of time. Thus, he makes the length of the test another factor to guide him in preparing the final form of the test.

During this phase of test construction, the author may see some resemblance to the game of chess. He has to consider factors of item discrimination, difficulty of questions, functioning of responses, curricular validity, balance of types of questions, and over-all length of the examination. Before making a move to adjust one factor, he has to consider the effect upon the other factors. He may find that several alternate adjustments appear possible, and he then has to decide which move will work out best in the end.

Although authors differ somewhat in the order in which they deal with factors for revising the experimental test, most of them follow the same general pattern. In a typical arrangement of steps, Phillips (36) first eliminated tests that lacked discrimination, then made adjustments needed to retain the correct proportion of questions shown in the table of specifications (curricular validity). Thirdly, she made adjustments on the basis of the criterion of difficulty of the question. Phillips found that while working on the curricular validity, she could simultaneously make adjustments to achieve a near 50-50 balance between true-false and multiple choice questions. Using these procedures, Phillips reduced the original test of 178 items to a final form of 100 items.

As an alternative to discarding unacceptable questions, the author may revise some of the items. However, he must expect to revalidate these items, which means administering the test a second time. When he makes a number of such item revisions, the author can expect to increase the item validity of the final form as compared with the original test. Broer and Miller (5) revised and restated questions which rated low in discrimination. They found that whereas their preliminary test contained approximately 30 items with a satisfactory discrimination index, the revised form of this test had 70 items with a satisfactory discrimination index.

TESTING FOR RELIABILITY

Reliability is of secondary importance in evaluating a written test, in contrast to the large part it plays in evaluating motor tests. When a written test meets the standards for discrimination, difficulty rating, curricular validity, and functioning of responses, it follows reasonably well that the test will be reliable. For this reason, the author need not test the reliability during the development of the test, although he may want to show a reliability coefficient as part of the over-all evidence in support of the test.

Although there are many procedures for testing reliability, they all fall into three major categories: (a) internal consistency reliability, (b) alternate forms reliability, and (c) test-retest reliability. In the fields of health, physical education, and recreation, the Kuder-Richardson procedure is the most often used technique for testing internal consistency and has been used in such studies as Langston (25), Dzenowagis and Irwin (8), Miller (31), and Phillips (36). The split-halves correlation, plus the Spearman-Brown formula, is used almost exclusively to compute alternate forms reliability. Fox (13), Waglow and Rehling (46), Stradtman and Cureton (42), Waglow and Stephens (47), Kelly and Brown (23), and Broer and Miller (5) have all used the split-halves technique. Because students would tend to remember the answers when retaking a written test, the test-retest reliability is of little use for knowledge tests.

Kuder-Richardson Procedure. The Kuder-Richardson procedure (37, 38) requires only one administration of a single test. Different formulas are used depending upon the assumptions that are made. One formula is used when the investigator assumes that the test measures only one factor. Another formula is used when, in addition to this previous assumption, it is also assumed that all intercorrelations between items are equal. By adding a third assumption—that all items have the same difficulty—a still simpler formula is permissible. Seldom do tests justify all three assumptions. However, the latter formula, which is simpler and quicker, may be satisfactory for some purposes. The authors state concerning this formula that, "It may be considered as a foot-rule method of setting the lower limit of the reliability coefficient, or the upper limit of error." It usually gives an underestimate, which is on the side of safety.

Split-Halves Procedure. Again, using one administration of a single test, the investigator correlates the sum of the odd-numbered questions with the sum of the even-numbered questions. This provides a Pearson r. Because this coefficient is based upon one-half the length of the test, the correlation is corrected using the Spearman-Brown prophecy formula (16) to get an estimate of reliability for the full-length test.

Split-halves reliability has the limitation that both halves of the test are taken at the same sitting. Therefore, if any chance factors exist that would tend to affect test performance differently at different sittings, the effect of these factors would be missing in scores collected in one sitting. Putting it another way, the chance factors would affect the split scores in the same direction. The over-all effect is to raise the correlation coefficient spuriously.

Angoff's Equation. Angoff (2) proposes a formula as a substitute for the Kuder-Richardson procedure. He warns that, although the Kuder-Richardson formula is intended for use in a single application of a single test, it assumes a correlation between separate forms. Hennis (17) used Angoff's Equation C to test the reliability of seven knowledge tests.

Guilford (16) points out that the following factors affect the reliability of a test:

1. *Item difficulty.* Items of moderate difficulty, where 50 percent pass and 50 percent fail, are the most reliable.
2. *Item intercorrelations.* Reliability is highest when the items of the test all intercorrelate highly.
3. *Range of difficulty.* The more nearly equal are the difficulties of the test items, the higher is the test reliability.
4. *Length of test.* Reliability increases with an increase in length of the test.
5. *Item discrimination.* Item discrimination, which is the correlation of an item with the total test score, is a good index of item intercorrelations (see No. 2, above). An effective way to increase the item intercorrelations is to improve the discriminative quality of the test items.

Ebel (9) reminds us that reliability values also are influenced by the types of subjects used in developing the test. He points out that it is easier to get high reliability when the students range widely in level of achievement than when they are more nearly equal.

The reader will notice that the above factors bear a close relationship to the factors that affect the validity of the test, as described above in this chapter. This similarity in factors affecting test validity and reliability illustrates why it can be said with confidence that when a written test has validity it may also be expected to be reliable.

PREPARING NORMS

Norms are tables that may be used to interpret test scores. The teacher can use these norms to tell whether the scores of his students are average, above average, or below the expected level of ability. It is customary for the author of a standardized test to prepare norms to accompany the test. This requires that he first administer the test to a number of subjects (a sample) from the population for whom the test is intended, and then prepare the norm table using the data thus collected.

Drawing a proper sample requires technical knowledge, and the reader is referred to Chapter 4, "Populations and Samples," for hints on how to draw a sample of subjects. It should be pointed out that the test author may be skilled in sampling procedures and still go astray at another point, namely in identifying the appropriate population from which to draw the sample. For example, we may expect that a group of students who have had no instruction in an activity will score lower on a test than a group that is comparable except for the fact that it has had one year of instruction in that activity. Therefore, if the test author administers the test to subjects at the beginning of the semester, the results will yield different norms than would result from year-end testing.

The point here is clear. The test author must first decide what kind of norms he wants, and then must sample from the kind of population that will produce these norms. Langston (25) wanted volleyball knowledge norms that reflected the results of volleyball instructions; therefore, he administered his test to students who had completed a course in volleyball instruction. Phillips (36) found that the badminton norm data she collected fell into two categories; some of the students had participated in 12 to 16 badminton classes while the others had been in 25 to 36 classes. She wisely set up two norm tables, and labeled one for beginners and the other for intermediates.

When the norm data are collected, the author is ready to construct the norms. Sometimes, norms are presented simply as a letter grade scale based on equal intervals on the baseline of a normal curve. The following norms, based on a distribution mean of 70 and a standard deviation of 10, illustrate this type:

A	88 & up
B	76-87
C	64-75
D	52-63
F	51 & down

Scales that range from 0 to 100 are frequently used, possibly because teachers and students are accustomed to this type of scoring on report cards and in classroom grading procedures. Sometimes this is a percentile scale, as in Phillips (36), although more frequently a standard scoring scale provides the 0 to 100 range.

Two commonly used forms of standard scoring scales are the T-scale and the sigma scale. The mean is 50 for each, but the T-scale has a standard deviation of 10 whereas the sigma scale has a standard deviation of 16.67. In comparison, these two scales have advantages and disadvantages, although the sigma scale appears to be more widely used in health and physical education tests. The main advantages of the T-scale are its ability to differentiate the achievements of the outstanding performers who tend to cluster at 100 when scored on the sigma scale, and the recognition of the poorer students' abilities by a value somewhat greater than zero. However, the T-scale has the disadvantage of usually placing 100 out of reach of the best performers in most classes, thus possibly tending to dampen motivation.

Physical Performance Tests

As stated previously, most of the physical performance tests are in physical education and recreation, in contrast to health education in which there are none except tests of proficiency in first aid. Textbooks generally agree that these tests fall into at least five major divisions, including anthropometric and body mechanics, cardiorespiratory, physical fitness, general motor ability, and sports ability tests. In turn, some of these are subdivided. For example anthropometric and body mechanics tests may include tests of posture, some tests of nutritional status, and tests of body build. Physical fitness has many components, and tests exist to measure some of these specific aspects. There are strength tests, flexibility tests, and endurance tests, to name a few. For a more complete breakdown of types of physical performance tests, the reader may consult a number of fine textbooks on measurement and evaluation in the fields of physical education and recreation.

In general, the method of constructing tests is the same for all types of physical performance tests. Furthermore, these test construction procedures are quite similar to those for written tests. They will be presented here in four parts: selecting the criterion, selecting the test items, testing the reliability and objectivity of the test items, and validating the test. A fifth step, preparing norms, has been covered previously under written tests, and the same procedures apply to physical performance tests.

SELECTING THE CRITERION

Sometimes, in developing tests, test authors select tentative test items as the first step, while at other times the investigator turns first to the problem of choosing the test criterion. The order in which the various steps occur depends somewhat upon the way in which the test is validated. Simply as a convenience, the selection of the criterion is presented first.

A criterion is a known and accepted measure of whatever the author wishes to test. In developing a test, the author selects a criterion, chooses an experimental test, and then correlates the test against the criterion. If the correlation is high, this may be interpreted to mean that the test accurately measures the criterion. If the criterion is a measure of physical fitness, then a test that correlates high with the criterion is also a measure of physical fitness. If the criterion is a measure of motor ability, then the test that correlates high with the criterion is also a measure of motor ability. We can see, then, that the criterion is the means used to validate the test. The *proven* validity of a test depends upon the extent of its correlation with a criterion. Several types of criteria are used in test construction. Some of the more valuable types are described briefly here.

Previously Validated Tests. When tests are known to be valid, they may be used as criteria against which to validate experimental tests. Previously validated tests are available in some areas and not in others. As test development progresses, more tests will become available for use as criteria. Wilson (51) used the Rogers' Short Strength Index, 3 x sum of right and left grips plus (pull-ups + push-ups) $\times \left(\frac{\text{weight}}{10} + \text{height} - 60\right)$, against which to validate her tests of strength. Because the criterion measures arm and shoulder girdle strength, the Wilson tests may also be said to measure arm and shoulder girdle strength. Phillips (35) validated his three-item test (vertical jump, chins, and running) against the Larson Muscular Strength Test (chins, dips, and vertical jump).

Competitive Standings. This technique is used mostly as a criterion for a sports skill and generally takes the form of a round robin tournament. To illustrate, Miller (30) administered a round robin badminton tournament to 20 players and then correlated the standings in the tournament with scores on an experimental badminton wall volley test. Using the tournament as a criterion of total playing ability in badminton, Miller concluded that the volley test validly measures ability in this sport.

Subjective Ratings. Judges' ratings can be a satisfactory criterion if the judges are competent and well trained and if they have an adequate chance to observe before rating. Judges' ratings are probably the most common criterion for validating tests, and the following examples are only a few of many presented in the literature. Everett (11) used a coach's ratings of baseball playing ability as the criterion against which he validated his experimental baseball test. Fox (14) validated a swimming power test against the criterion of judgment-of-form swimming. The judge was an experienced teacher of swimming who judged each subject on a 10-point scale according to form in the sidestroke and front crawl stroke.

One criticism of the judges' ratings is that it introduces the element of subjectivity. However, this element may be largely overcome by having the judges use a checklist containing the factors to be judged and a scale for standardizing the ratings.

Divergent Groups. Another criterion for validating tests is the divergent groups procedure. This criterion may be used any time two groups can be found to represent the opposite extremes of the quality that the test is to measure. Thus, in the development of the McCurdy-Larson Test of Organic Efficiency (29), varsity swimmers and infirmary patients represented good and poor organic efficiency, respectively. In developing a motor ability test for high school girls, Kammeyer (20) used girls who had participated constantly in the extracurricular athletic program and girls who participated very little in the same program as divergent criterion groups representing high and low athletic ability, respectively. In each of these studies, the authors found that subjects in the "good" group scored higher on the test than did the subjects in the "poor" group, thus providing evidence of test validity.

Composite of Criterion Factors. Sometimes, a test author is not satisfied with any single test as a criterion, possibly because no one test contains all the factors he wants in his criterion measure. In this situation, the experimenter may combine several tests or test elements to gain a composite of the factors he wants. For example, Barrow (4), looking for a criterion of motor ability, combed the literature to find 29 test items representing the factors of agility, hand-eye and foot-eye co-ordination, power, speed, arm and shoulder co-ordination, strength, balance, and flexibility. He used this composite as the criterion against which he validated two test batteries for predicting general motor ability for college men.

Some investigators have resorted to the technique of factor analysis to find the factors for their composite criterion. For example, Larson (26) factor-analyzed 16 test items and found two factors, namely dynamic strength and static dynamometrical strength. He then used the factor of dynamic strength as a criterion against which he validated the Larson Strength Test.

Descriptive Criterion. Some authors have validated their tests by using narration to show that the characteristics of the test are similar to the qualities that the test is supposed to measure. Here, the criterion is a description or definition of the quality to be tested, and the investigator validates his test against this descriptive criterion by the process of logical explanation. This is a form of self-validity, sometimes referred to as face validity. Weiss (49: 48), who used descriptive criteria to develop tests in 19 skills of softball, football, and soccer, put it this way:

> "Each testing procedure was devised in such a manner that its use involved performance in the skill itself. Since it was this same skills performance that was being evaluated, the criterion and the testing device became one and the same, and validity was considered to be inherent within the device."

Sometimes test constructors use more than one criterion when validating their tests. To illustrate, Kammeyer (20) first validated a motor ability test against a seven-item athletic achievement composite criterion, and then validated the test against the criterion of two groups which differed in extent of athletic participation.

The literature reveals that test authors do not favor any one type of criterion over others. Rather, there is widespread use of all of the criteria described above, sometimes modified to fit a particular situation. However, constructors are careful to select validity criteria that they can defend, for they know that a test is no better than the criterion against which it is validated.

SELECTING THE TEST ITEMS

Equally as important as the choice of the validity criterion is the choice of the experimental test items. Although some authors have developed acceptable tests after selecting test items through trial and error, the choice of test items by rational and considered judgment is a much more satisfactory procedure. Test constructors use any or all of the following criteria to guide them in choosing experimental test items.

Relationship to the Criterion. Because the final test will be validated against the criterion, it is logical for the test constructor to select test items that appear to be the best estimate of the criterion. The following experiences illustrate this point.

Miller (30), wanting a test to predict total performance in badminton, used a round robin tournament as a criterion of badminton playing ability. Looking for the best estimate of this playing ability, Miller found that the clear was used more than any other stroke during the doubles and singles contests of a national badminton tournament. By experimenting, with the aid of photography, Miller found that the ideal-driven clear could cross the net as low as 7½ feet above the floor and still go over an averagesized opponent whose racket was outstretched. Using this information, she devised a wall volley test in which the subject volleyed the bird continuously against a wall above a line $7\frac{1}{2}$ feet high and scored on the basis of the number of fair hits within a 30-second time period. A validity coefficient of .83 between this test and the criterion rewarded this test constructor's logical approach to the selection of an experimental test item.

As mentioned previously, Barrow (4) used a 29-item battery as a criterion of general motor ability, wherein the 29 items represented eight factors (see page 240) essential to motor ability. In selecting test items to predict this criterion battery, Barrow saw no reason to go beyond the 29 criterion items. Within these 29 items, he found a six-item combination which correlated .95 with the criterion. Further, he found that three items of the six-item test combined to predict the criterion with a correlation of .92. This method of building a test from items within the criterion battery has been used by several test authors but does yield a spuriously high coefficient.

Reliability and Objectivity. In selecting experimental test items, the test constructor should give preference to items whose reliability and objectivity are known. When an investigator selects unreliable items, he condemns his efforts at that point, for it is well known that an unreliable test cannot be valid. The reliabilities of many items have been published during the past decade, and the wise experimenter makes use of this information, whenever he can, in selecting items with which to experiment. More is presented about reliability and objectivity below in this chapter.

Independent Performance. Each test item should measure one performer, and his performance should not be affected by another person, either student or test assistant. For this reason, it is harder to test certain abilities than others. For example, anyone devising an objective instrument to measure boxing ability would have to find some way to eliminate the influence of the opponent. A person may look good against one opponent, and at once look bad against a better boxer.

Realism. When constructing sports skills tests, the investigator should try to use items which are like the game situation. For example, in measuring a football pass for accuracy, it would be better to have the subject pass to a moving rather than a stationary target since the moving target situation occurs more often.

Scoring. Objective scoring systems usually are more reliable than subjective scoring. For this reason, the investigator should select an objective method for scoring a test item, if there are alternatives. For example, counting the number of baskets made would be an objective method of scoring basketball shooting ability whereas rating the player's shooting form would be subjective. Or, if a choice is to be made between two test items of comparable quality, the choice might well be the one with the more objective scoring procedure. In addition, the scoring system as well as the nature of the test should provide a normal distribution of scores.

Practicality. Other things being equal, the investigator should select the test items which are most practical to administer. Amount of equipment needed, time required to administer, space requirements, administrative ease, and leadership requirements are some of the factors that affect the usefulness of the test.

Suitability. The test constructor should know that a test item can appear to be related to the criterion and yet not be suited to the test he wants to develop. As an obvious example, chinning is considered to be a valid measure of muscular strength and endurance, but it could not be used in testing kindergarten boys since it demands more strength than they possess. Likewise, some excellent items in boys tests would not be suitable for testing girls. The test constructor should be certain that the test item suits the group for whom the test is intended.

RELIABILITY AND OBJECTIVITY

Reliability is the consistency with which a test can be administered by the same tester. Reliability can be influenced by such extraneous factors as the time of day, the equipment, momentary attitude of the subject, conditions in the surrounding area, such as heat, light, and humidity, and lack of specific directions for performing the test. When these and any other extraneous factors are controlled, reliability improves.

Objectivity is the consistency with which a test can be administered to the same subject by different testers. Objectivity is influenced by the judgment of the tester. Items that can be scored with the least need for judgment have the most objectivity. To illustrate, we could expect a basketball shooting test item to have better objectivity than a posture rating test because the tester will find it easier to count the number of baskets scored than to judge body alignment.

The usual method of finding test reliability is to administer the test to the same group twice and then compute the reliability correlation coefficient. Objectivity can be tested at the same time by having two testers score the subjects independently during the first test period. A correlation between the two sets of scores in that first period will yield an objectivity coefficient. It should be noted that extraneous factors that affect reliability do not influence objectivity when two testers score performance at the same time. The obvious reason is that they both score the subject under precisely the same conditions and only their own judgment will cause them to differ in scoring. However, if one tester tests a group and at another time a second tester tests the same group, the resulting test-retest correlation coefficient is a combination of reliability and objectivity. Such a coefficient is hard to interpret, as one cannot be certain what proportion of inconsistency results from the testers' judgments and what part is due to other extraneous factors. Nevertheless, this combined coefficient is probably a more realistic evaluation of test administration consistency than either reliability or objectivity separately.

When a test item calls for several trials, its reliability may be found with a single administration. To illustrate, if 10 trials are given in a baseball throw for accuracy, the sum of the odd-numbered throws may be correlated with the sum of the even-numbered throws. This gives a correlation for 5 trials. To estimate the reliability for 10 trials, the correlation obtained on half the test is stepped up by the Spearman-Brown Prophecy formula.

The generally accepted standard of test reliability is .85 for individual use, and .75 when the test results are used to evaluate group achievement. Different levels are set by some test constructors. Certain types of tests, especially measures of accuracy, are known to have low reliabilities, and many investigators have set .70 as the minimum acceptable reliability for these tests.

Although we are inclined to think of reliability as an inherent test characteristic, the test constructor should remember that he can help make the test yield its potential reliability by preparing a complete set of instructions for administering the test. Guided by these instructions, the subject is more likely to perform the test as intended. Under these standardized conditions, if the test item reliability is still low, the item should be discarded, or the number of trials increased, or the test otherwise remodeled. Correlations should be repeated after each successive revision and the process continued until the test is found to be satisfactory. In revising the test, the Spearman-Brown Prophecy formula may be used to estimate the number of trials needed to obtain a desired level of reliability.

Test authors should test the reliability of each test item, whether or not the reliability of the item has been previously published. His research presents a set of conditions which may differ from other studies, and reliabilities will change accordingly. For example, one investigator can obtain a higher reliability coefficient than another simply by administering the test item to a group with a wider range of performance in that item. Adkins (1:157-58) discusses this element of variability in connection with reliability, as well as many other aspects of the problem of reliability of test scores.

When a total score is computed as a composite of two or more test items, the test author should also demonstrate the reliability of this total score.

VALIDATING THE TEST

A test is valid when it correlates high with the test criterion. In developing a test, the test constructor should demonstrate the validity of his instrument. Two validating procedures are used—descriptive and statistical.

Descriptive Validity. An investigator validates a test descriptively when he shows by logical explanation that the test does what the descriptive criterion calls for. The technique is logical explanation. The ingredients are a description of the test and a description or definition of the quality to be tested. The validating process calls for the investigator to reason convincingly that the test is indeed a measure of what is defined as the criterion. This reasoning, presented narratively, is offered as evidence of test validity. The investigator should document his statements when he feels they might be challenged.

To illustrate the process of descriptive validity, let us consider what an investigator would have to say about the 50-yard dash as a measure of speed in running. First, he would explain running speed. Because speed can be interpreted several ways, he would have to explain the meaning of speed in precise terms. If he is interested in *pure* speed without endurance, he would contrast running speed with endurance and speed unaffected by endurance. Through this description, the concept of continuous top speed without dropoff owing to fatigue should emerge. Having first presented this descriptive criterion of running speed, the investigator next describes the 50-yard dash. If he presents this description in terms that demonstrate this event to be a measure of the criterion, he can use this description as evidence of descriptive validity. Certainly, he would point out that in the 50-yard dash the runner accelerates to top speed as quickly as possible and that he tries to maintain this speed the full distance. At this point, if he could state that it is possible for the average person to maintain top speed for 50 yards, he would strengthen the validity of the event as a measure of speed unaffected by endurance. If he could cite evidence in the literature that 50 yards can be run at top speed without dropoff, so much the better.

As mentioned previously, descriptive validity is a form of self-validity and is often called face validity. Fox (14) assumed face validity for her swimming power test. Moyna (34) used face validity to validate a group of test items to measure motor performance. Although investigators continue to validate tests descriptively, some consider this technique to be subjective and inconclusive. They feel that conclusive evidence of test validity must be demonstrated by statistical validation.

Statistical Validity. When an investigator uses any of several statistical formulas to correlate a proposed test against the test criterion, he validates the test statistically. Statistical validity procedures can range from simple to complex, depending upon the size of the test and the degree of the investigator's thoroughness.

Perhaps the simplest validation is a correlation between a single-item test and the test criterion. This requires no more than a simple product-moment correlation if numerical data for the test and the test criterion are both measured on a continuous scale. Other correlation techniques can be used when either or both variables are not continuous. (See Chapter 7 for descriptions of various statistical procedures mentioned in this chapter.) Miller (30) found that her badminton wall volley test had a validity coefficient of .83 when correlated against the criterion of a round robin tournament. Mohr and Haverstick (32) obtained a validity coefficient of .79 when they correlated a wall volleying test of volleyball ability against the criterion of judges' ratings.

When the investigator constructs a test of two or more test items, he can validate the test by using the multiple correlation and regression equation techniques. The multiple correlation indicates the degree of accuracy with which the battery of test items can be used to predict the criterion measure. The regression equation shows how much each item contributes to the prediction. The test is validated in four steps.

Correlations with Criterion. First, the investigator computes a correlation between each test item and the criterion. Those items which correlate high with the criterion are the best prospects for the new test. However, the test author does not need to make a choice of test items, as yet.

Intercorrelations. Next, the investigator correlates each test item with every other test item in the experimental battery. Called intercorrelations, these coefficients are computed in the same manner as in the first step. These intercorrelations are the basic ingredients of the next (third) step along with the correlations in the preceding step. The effect of the intercorrelation is to eliminate duplication of measures. If two items each correlate high with a criterion and also interrelate high with each other, one can be discarded as unnecessary. The high intercorrelation means that the two items combined will be no better as a test than a single-item test, using either item. In fact, the two-item test would have less value since it would take more time to administer.

Multiple Correlations. In the third step the investigator computes the multiple correlations to express the degree of relationship between the criterion measure and two or more test items. Here, the Wherry-Doolittle method of computing multiple correlations is a time-saver, since it allows the computation of multiple correlations directly from the correlations and intercorrelations previously described. This is a great improvement over earlier techniques which required a person to compute partial correlations and partial standard deviations. As the number of test items increases, the mechanics of computation by the older methods become prohibitive.

The Wherry-Doolittle method has other advantages. With it, the investigator can select the test items analytically and compute the effect upon the multiple correlation of adding them one at a time. When the multiple correlation does not increase, the optimum number of items is reached.

Sometimes the investigator must consider other factors than validity in making the final choice of the test battery. He may want to substitute an item of lesser validity (not too much less) in order to gain a test which is less time-consuming or which requires less equipment. Whatever the choice of test items, the investigator will demonstrate test validity by the size of the multiple correlation.

Regression Equation. In the final step, the investigator computes the regression equation using the Wherry-Doolittle of the multiple correlation. The regression equation indicates the relative importance of each item in the test battery. If the items are all of approximately equal weight (importance), the test author can disregard weighting in setting up the scoring system. In this case, the total test score can be computed by converting test item scores to standard scores and then summing. However, if the weightings in the regression equation are unequal, it is best to use the regression equation to compute total performance score. If the equation is not used when the test items have unequal weight, the total test scores will be less valid than is indicated by the multiple correlation coefficient.

Both Barrow (4) and Everett (11) used the above four steps to validate their tests of motor ability and baseball ability, respectively.

There is not room here to describe all the modifications of statistical validation which test authors have used. However, they are all based on the same idea of computing some type of correlation between the test item or items and the criterion measure.

Once the final form of the test is selected and validated, the test author may want to construct norms for use with the test. The subject of norms was discussed earlier in this chapter.

To be useful, the newly developed test should be published with complete information of two kinds. First, the author should describe in detail the procedures he followed in developing the test, and should present evidence of test reliability, objectivity, and validity. Secondly, the author should give complete instructions for administering the test. This information may include purpose of the test, for whom intended (sex and age level), test items and equipment, leadership requirements, time requirements and numbers that can be tested, space requirements, organization of subjects, instructions to the subjects and to the leaders, scoring instructions and sample score forms, and norms, if any.

SUMMARY

Although this chapter presents certain fundamental steps which are essential in test construction, there are several acceptable ways to accomplish each of these steps. The nature of the test and the judgment of the investigator will help to determine the choice of techniques. For more detailed information about certain aspects of test construction, statistics, and test use, the reader should consult other references, including those in the bibliography at the end of this chapter. Other chapters in this book should be helpful also.

BIBLIOGRAPHY

1. Adkins, D. C. *Construction and Analysis of Achievement Tests.* U. S. Office of Education, Federal Security Agency. Washington, D. C.: Superintendent of Documents, Government Printing Office, 1947.
2. Angoff, W. H. "Test Reliability and Effective Test Length." *Psychometrika* 18: 1-14; 1953.
3. Aschenbrenner, R. A. *A Study of the Reliability of One Type of Discrimination Index for Test Items.* Master's thesis. Iowa City: State University of Iowa, 1949.
4. Barrow, H. H. "Test of Motor Ability for College Men." *Research Quarterly* 25: 253-60; October 1954.
5. Broer, M. R., and Miller, D. M. "Achievement Tests for Beginning and Intermediate Tennis." *Research Quarterly* 21: 303-13; October 1950.
6. Carr, M. G. "The Relationship Between Success in Physical Education and Selected Attitudes Expressed by High School Freshman Girls." *Research Quarterly* 16: 176-91; October 1945.
7. Davis, Frederick Barton. *Item-Analysis Data: Their Computation, Interpretation, and Use in Test Construction.* Harvard Education Papers #2, Graduate School of Education, Harvard University, 1946. 42 p.
8. Dzenowagis, J. G., and Irwin, L. W. "Prevalence of Certain Harmful Health and Safety Misconceptions Among Fifth- and Sixth-Grade Children." *Research Quarterly* 25: 150-63; May 1954.
9. Ebel, R. L. *How to Judge the Quality of a Classroom Test.* Technical Bulletin No. 7. Iowa City: The Examination Service of the State University of Iowa, September 1955.
10. Engelhart, M. S. "How Teachers Can Improve Their Tests." *Chicago Schools Journal* 25: 16; 1943.

11. EVERETT, P. W. "The Prediction of Baseball Ability." *Research Quarterly* 23: 15-19; March 1952.
12. FLANAGAN, J. C. "General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from Data at the Tails of the Distribution." *Journal of Educational Psychology* 30: 674; 1939.
13. FOX, K. "Beginning Badminton Written Examination." *Research Quarterly* 24: 135-46; May 1953.
14. FOX, M. G. "Swimming Power Test." *Research Quarterly* 28: 233-37; October 1957.
15. FRENCH, E. "The Construction of Knowledge Tests in Selected Professional Courses in Physical Education." *Research Quarterly* 14: 406; December 1943.
16. GUILFORD, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill Book Co., 1950. Chapter 17.
17. HENNIS G. M. "Construction of Knowledge Tests in Selected Physical Education Activities for College Women." *Research Quarterly* 27: 301-309; October 1956.
18. JOHNS, E. B., and JUHNKE, W. L. *Health Practice Inventory.* Stanford, Calif.: Stanford University, 1952.
19. JURGENSEN, C. E. "Table for Determining Phi Coefficients." *Psychometrika* 12: 17-29; 1947.
20. KAMMEYER, S. J. "Reliability and Validity of a Motor Ability Test for High School Girls." *Research Quarterly* 27: 310-15; October 1956.
21. KAPPES, E. E. "Inventory To Determine Attitudes of College Women Toward Physical Education and Student Services of the Physical Education Department." *Research Quarterly* 25: 429-38; December 1954.
22. KELLEY, T. L. "The Selection of Upper and Lower Groups for the Validation of Test Items." *Journal of Educational Psychology* 30: 17-24; 1939.
23. KELLY, E. D., and BROWN, J. E. "The Construction of a Field Hockey Test for Women Physical Education Majors." *Research Quarterly* 23: 322-29; October 1952.
24. KILANDER, H. F. *Kilander Health Knowledge Test for College Students.* Ann Arbor, Mich.: J. W. Edwards & Sons. 1959. 8 p.
25. LANGSTON, D. F. "Standardization of a Volleyball Knowledge Test for College Men Physical Education Majors." *Research Quarterly* 26: 60; March 1955.
26. LARSON, L. A. "A Factor and Validity Analysis of Strength Variables and Tests with a Test Combination of Chinning, Dipping, and Vertical Jump." *Research Quarterly* 11: 82-96; December 1940.
27. LIKERT, R. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22: 5-43; 1932.
28. McCUE, B. F. "Constructing an Instrument for Evaluating Attitudes Toward Intensive Competition in Team Games." *Research Quarterly* 24: 205-209; May 1953.
29. McCURDY, J. H., and LARSON, L. A. "Measurement of Organic Efficiency for the Prediction of Physical Condition." *Research Quarterly Supplement* 6: 11-41; May 1935.
30. MILLER, F. A. "A Badminton Wall Volley Test." *Research Quarterly* 22: 208-13; May 1951.
31. MILLER, W. K. "Achievement Levels in Tennis Knowledge and Skill for Women Physical Education Major Students." *Research Quarterly* 24: 81-90; March 1953.
32. MOHR, D. R., and HAVERSTICK, M. J. "Repeated Volley Tests for Women's Volleyball." *Research Quarterly* 26: 179-84; May 1955.
33. MOSIER, C. I.; MYERS, M. C.; and PRICE, H. G. "Suggestions for the Construction of Multiple Choice Test Items." *Educational Psychological Measurement* 5: 261; 1945.
34. MOTHS, K. E. *An Analysis of Badminton and Constitutional Body Patterns.* Doctor's thesis. New York: New York University, 1951.
35. PHILLIPS, B. E. "The JCR Test." *Research Quarterly* 18: 12-29; March 1947.

36. PHILLIPS, M. "Standardization of a Badminton Knowledge Test for College Women." *Research Quarterly* 17: 48-63; March 1946.
37. RICHARDSON, M. M., and KUDER, G. F. "The Calculation of Test Reliability Coefficients Based on the Method of Rational Equivalence." *Journal of Educational Psychology* 30: 681; 1939.
38. RICHARDSON, M. M., and KUDER, G. F. "The Theory of the Estimation of Test Reliability." *Psychometrika* 2: 151; 1937.
39. ROSS, C. C. *Measurement in Today's Schools*. Second edition. New York: Prentice-Hall, 1947.
40. SCOTT, M. G. "Achievement Examination for Elementary and Intermediate Swimming Classes." *Research Quarterly* 11: 100; May 1940.
41. SCOTT, M. G., and FRENCH, E. *Measurement and Evaluation in Physical Education*. Dubuque: Wm. C. Brown Co., 1959.
42. STRADTMAN, A. D., and CURETON, T. K. "A Physical Fitness Knowledge Test for Secondary School Boys and Girls." *Research Quarterly* 21: 53-57; March 1950.
43. SWINEFORD, F. "Validity of Test Items." *Journal of Educational Psychology* 27: 68; 1936.
44. THURSTONE, L. L., and CHAVE, E. J. *The Measurement of Attitude*. Chicago: University of Chicago Press, 1929.
45. VOTAW, D. F. "Graphical Determination of Probable Error in Validation of Test Items." *Journal of Educational Psychology* 24: 682-86; 1933.
46. WAGLOW, I. F., and REHLING, C. H. "A Golf Knowledge Test." *Research Quarterly* 24: 463-70; December 1953.
47. WAGLOW, I. F., and STEPHENS, F. "A Softball Knowledge Test." *Research Quarterly* 26: 234-43; May 1955.
48. WEAR, C. L. "The Evaluation of Attitudes Towards Physical Education as an Activity Course." *Research Quarterly* 22: 114-26; March 1951.
49. WEISS, R. A. *The Construction of Achievement Scales for the Measurement of Performance in Selected Physical Education Activity Skills*. Doctor's thesis. New York: New York University, 1949.
50. WEITZMAN, E., and MCNAMARA, W. T. "Apt Use of the Inept Choice in Multiple Choice Testing." *Journal of Educational Research* 39: 517; 1946.
51. WILSON, M. "A Study of Arm and Shoulder-Girdle Strength of College Women in Selected Tests." *Research Quarterly* 15: 258-67; October 1944.

CHAPTER 9

Descriptive Methods

ELWOOD CRAIG DAVIS
LAWRENCE RARICK

THE MAIN PURPOSE OF A RESEARCH STUDY MAY BE TO ENUMERATE or depict the characteristics, abilities, behavior, or opinions of subjects; to delineate through words or quantitative values the status of a group, institution, structure, or other facilities; or to portray the trends or changing values of the characteristics of human beings or objects. Research studies may be limited or broad, intensive or highly specific, brief or continuous and lengthy. Whatever the defined purpose of the study, the end result should be a verbal and statistical picture which has objective, clear details, is valid or true to fact, and shows proper perspective or interrelationships.

The methods described in this chapter are used for various kinds of research. The purpose of each method is explained.

The Survey

ELWOOD CRAIG DAVIS

The survey may be considered a research medium if it meets certain criteria. For example, it may use valid sources and pertinent, valid, reliable, and accurate methods, techniques, and tools—and thus yield acceptable data for the interpretative and generalizing processes. In many surveys it is possible to seek, find, and report all pertinent facts.

Nevertheless, the survey as a research medium is particularly susceptible to incomplete reporting of the pertinent facts. The *general purposes* of the survey are to reveal current conditions, to point up the acceptability of the status quo, and to show the need for changes. If the survey involves the help of several persons in addition to the surveyor and if they are not interested in conducting careful research or in carrying out the purposes of the survey, the pertinent facts remain either undiscovered or unreported. Furthermore, some surveys are financially supported by persons who, when they see the findings, resolutely object to a certain finding and demand that no report be made of it. Thus, regardless of the original intent of the surveyor and regardless of the quality of his preparation, circumstances beyond his control may prevent his survey from meeting some research criteria.

These unfortunate possibilities do not eliminate the survey as a method worthy of consideration in conducting some research. Nor do these possibilities mean that the survey may not help to bring about improvements in educational, civic, and other situations.

It becomes apparent from the foregoing that a survey purports to be an orderly collection, analysis, interpretation, and report of pertinent facts and information concerning an enterprise or situation or some aspect thereof, insofar as conditions and circumstances permit.

STEPS IN THE SURVEY

The ten titled paragraphs below outline the steps to follow in using the survey as a research medium.

Studying Situation and Problem. Each situation to which the survey is applied is unique. One phase of this uniqueness is the problem or problems creating the need for the survey. Other unique aspects of a situation may reflect local conditions and circumstances. Still others may be the limitations that modify or circumscribe the methods or tools to be employed. Finances and time are two examples.

Consequently, before determining the exact purposes of a given survey, the surveyor should consider the specifics underlying and connected with the local situation. In addition to those indicated above, there may be such factors as availability and reliability of pertinent records, nature and source of human assistance, climate and topography, nature of the issues underlying the problem(s), possible legal provisions, and the nature of the sources of information.

Formulating Purposes. Formulating the purposes of a given survey serves to indicate the foci of attention and effort. This step also indicates sources of information, methods and tools or devices for obtaining information, emphases to be made in the survey report, bases of interpretation, and the nature of the recommendations. As elsewhere, goals or purposes not only are of primary consideration, but they also form a powerful force in determining the nature of other and related steps in research. The surveyor also should be aware that purposes arise from basic beliefs about values, relationships, meanings, and other *philosophic* matters. The formulation of the exact purposes of a given survey should begin long before the survey begins. The basic beliefs of many persons may be involved in this task. The resolving of conflicting basic beliefs is seldom accomplished hastily.

Considering Type, Scope, and Nature. Once the situation and its attendant problem or problems have been studied and the purposes of the survey formulated, the surveyor is ready to plan his next steps. Outlining of the plan should follow the general structure of any well-planned research. Some surveyors go a step further and prepare a topical outline. This is a preliminary and tentative plan of the topics and subtopics and their sequence as they may appear in the final survey report.

During this planning stage, the surveyor carefully notes and labels the exact nature, scope, and type of the survey. All limiting, supporting, and modifying forces and factors are again considered and related to the specific purposes of the survey.

Keeping in mind that the survey may be an appropriate way of attempting to solve certain local problems in any one of the three related fields of health, physical education, and recreation and their subdivisions, there are other aspects of the scope, nature, and type of survey which must be considered in building a plan of work. For example, a surveyor may need to find out about local ordinances, charters, rulings, codes, laws, policies, procedures, and similar provisions which may relate to legal, quasi-legal, and regulative duties, obligations, rights, and limitations of authority applied to organizations and persons with whom he deals.

Further, he may find it necessary and desirable to obtain information about the nature of the local community. The following items could be considered: population trends, mobility, size, and concentrations; types of occupations represented in the community; sociological and economic strata and groupings; social conditions with reference to such basic considerations as health and delinquency; vital statistics; religious groupings; adequacy of medical care and hospitalization; the public relations of the situation being surveyed; sources of food and water; and similar concerns.

There is still another area determining the nature, scope, and type of survey which the surveyor should investigate during the planning stage. This third area may involve the following items of information: relationships of the situation being surveyed to pertinent councils, commissions, boards, committees, associations, and other organizations whose authority or position affects, or is related to, the operation and structure of the situation, as well as to some aspect of the survey.

A fourth area of preliminary consideration is finance. Examples of this item are: satisfactory financial evidence that the survey can be completed, including the production and possible publication of the survey report; assets and liabilities of the situation in regard to ability to pay for *probable* recommended changes; bases of taxation; financial history of the enterprise; any financial commitments and probable expenditures competing with demands for money to pay for the carrying out of the survey's recommendations; restrictions as to bond issues; and the entire budget of the situation.

A fifth area of preliminary study is that of personnel. This item refers to each person who is to assist in any way in conducting the survey. It also refers to each person who is to serve directly or indirectly as a source of information. Consideration is given not only to the number of these persons but to their adeptness in the functions they are to perform and to their reliability, availability, and capacity to serve in the ways planned.

A sixth area to be studied consists of supplies, facilities, and equipment. How many different locations (building, fields, plants, centers, etc.) are to be surveyed? What are the traveling time and the distance between them? Are records, minutes, and other pertinent written and printed matter, such as results of tests, examinations, and programs, readily available? Are all supplies, equipment, and facilities to be used by the survey staff adequate, available, and ready? Is there any reliable evidence from other studies that may supplement the observation of the surveyor and his evaluation of facilities, equipment, and supplies?

Securing Co-operation. The survey at its best, like any research medium, demands the co-operation of a number of persons. If the findings are to be reliable from a research viewpoint and if they are to have the support of those for whom the survey is conducted, the co-operation of the survey staff with the local personnel is essential. Considerable preliminary work with local persons is necessary in securing their full co-operation. In this advance briefing, the surveyor must present to local authorities the general kinds, amounts, and nature of the data and information, together with the methods of collecting, that will be used in the survey. Such a step also helps to foster co-operation between local authorities and their staffs, without which the survey is apt to prove ineffective.

Such considerations as these suggest that the survey should be conducted both by outside specialists and by local specialists who have the respect and confidence of their peers and their superiors. If the survey is to accomplish its over-all purpose of improvement or of justification of status quo, there must be belief in what is done and what is found, so that follow-through is desired, effected, and carried on by the local personnel.

Selecting Participating Personnel. The trend is for general survey staffs to include specialists in the three related fields, when expertness in one or all of these fields adds authenticity, reliability, and validity to the survey findings and recommendations. These specialists also should be adept in one or several of the techniques employed in their parts of the over-all survey, such as the interview, testing, and questionnaire. If a choice must be made between a specialist in a given field and a person expert in some survey technique, experience has shown that it is far better to choose the specialist in a professional field and instruct him in the use of survey techniques rather than to attempt to school the expert in survey techniques in the significant content and methods of a professional field.

As indicated above, local members of the survey team should be selected who not only are co-operative but who are also well prepared professionally and respected by their local colleagues. The precise functions to be performed by these local specialists should be established as early as feasible, so that instruction in appropriate survey techniques may be given. The vast increase of persons in the three related fields who have earned, or who are studying for, advanced degrees indicates an increase in the number of those prepared to use at least one of the survey methods or techniques.

Finding Sources of Data. Unlike some research media, the survey often uses a wide variety of major sources of information. These major sources of data are *documentary* (records, reports, films, any printed materials); *functioning of processes* (teaching, administration, supervision, coaching); *human* (pupils, teachers, principals): *facilities, equipment, and supplies;* and *natural elements* (topography, climate, soil, water). (8)

Obviously, the information from some of these sources is "at hand" before the survey begins. The surveyor should make sure of the accuracy, authenticity, and suitability of all these sources. During the survey, other information is collected from sources by members of the staff, under the supervision of the surveyor.

Collecting Data. Here again the survey is characterized by a variety of techniques. Multiplicity of types of sources usually means variety in techniques. Here are some commonly used methods of gathering survey information: observation, study of documentary data, interview, score card, tests, inspection, examinations, job analysis, case study, tape-recordings, photography and movies, and the questionnaire (8). Some of these methods may appear to the novice to be sources of data, for example, the questionnaire and tape-recordings. Such ways of obtaining data serve as sources *after* the data are gathered, but these particular data come from *human* sources.

No matter how carefully a collecting process may be constructed and used, the data gathered may be suspect if the source is suspect. Similarly, no matter how acceptable the source, if the emerging data are obtained inaccurately, their use is contraindicated.

As in so many aspects of the survey, sources and collecting methods must be geared to specific purposes of the survey, local conditions and circumstances, time and money available, the adeptness of staff members in using survey techniques, and similar conditioning and limiting factors.

Interpreting Data. A key step in the survey, as in all efforts to find dependable information, is the interpretation of the findings. Some research experts refer to this step as the "acid" test of the quality of the research effort. Certainly, care in selecting and using acceptable sources and methods of collecting data is fundamental but it also is *preliminary* to seeing and pointing out the meaning, significance, and pertinency of the data.

Examples of aids to interpretation (in addition to having dependable, adequate data) are such processes as (a) *classification* of the data into meaningful categories and (b) *statistical treatment* of the data. It also helps to know the mental operations involved in interpretation, such as (c) *establishing hypotheses,* (d) *drawing inferences,* (e) *making judgments,* (f) *predicting,* (g) *reasoning constructively,* (h) *making assumptions,* (i) *drawing analogies,* and (j) *being* intellectually *discriminating.* In short, one uses these mental operations in attempting to accurately answer the question, "What do these data mean?" Still further understanding of the interpretative process comes from considering what the surveyor does with the data in order to get at its meaning. He (k) ferrets out, searches for, and tries to sense, to see, and then to show *relationships* among the data. He (l) *synthesizes* or groups the data into the largest assimilable whole. He (m) may *compare* one finding with another or with the whole or with some "outside" fact, standard, or criterion. He (n) may *analyze* one datum or several data, or the data as a whole. Analysis here refers to the breaking-down, the resolving, the determining of the elements (the constituent smaller parts) of some given finding. He (o) also may simply make a reference to some phenomenon, some factor, some occurrence as a way of helping to show what the data mean.

Nineteen devices were found to be operative in using ways (a) to (e) of interpreting data mentioned above. These devices (8) are listed here in order of their preference, as determined by four juries of raters, which included experts in research: acceptable scientific standards, expert agreement, accepted standards, tests, external comparison (e.g., findings in another situation), accepted studies, graphical techniques(charts, tables), expert opinion, statistical data, descriptive factual materials (such as another survey report), internal comparison, common sense judgment, group opinion, photographs, prevailing practice, existing conditions, surveyor's opinion, hypothetical criteria, and someone's opinion.

One of the most easily overlooked, yet one of the most important, aspects of interpretation is the viewing of the data against the backdrop of large ideas and large concerns—a frame of reference that transcends the concerns of the three related fields. Such backdrops are needed to see and point up meanings of the data in order to maintain perspective.

Another phase of the interpretative process frequently overlooked is the desirability of consistency in holding to the fundamental viewpoint assumed by the surveyor. In a survey, this "home-base" is usually a consideration of social implications, values, and involvements as distinguished from the personal or individualistic view. The well-trained surveyor should write his interpretation from one over-all, all-pervading viewpoint, even though within this panorama some recognition of more immediate, narrower viewpoints may be indicated.

Still another factor that the novice surveyor may be unaware of is the taking of positions or drawing of inferences which are not supported by the data. A person may unconsciously allow his personal wishes and basic beliefs to influence the interpretative process. Interpretation, a subjective process at best, is error-prone when the surveyor is not alert constantly to such possibilities.

Another warning to the novice surveyor seems advisable. There is a human tendency to give undue importance to numbers—to unthinkingly assume that some superiority is automatically attached to the larger of two numbers and to ignore the smaller number. As important and desirable as it may be to have data in numerical form when possible, and to treat them with *suitable* statistical operations, these operations solve no situational problems. The more thoroughly familiar a professionally prepared person is with statistics, with the assumptions and hypotheses upon which they are based, and with the practical application and follow-through of statistics, the more cautiously he deals with numbers. This warning does not mean that suitable statistical treatment of survey data is to be used less often. Actually, it is a warning that the surveyor should be, or become, thoroughly adept in statistics, its uses, and its basic assumptions to ensure better use of this aid to interpretation.

When the survey is regarded as research, conclusions should be drawn and labeled as such. Conclusions go beyond a statement of the major findings given in concentrated, general terms; they are generalizations *based on* the major findings. Conclusions include the implications of such findings and aid the reader in seeing how and where the major findings fit into the larger picture.

Recommendations based on conclusions as well as on major findings have the advantage of reflecting the long broad view and the deeper meaning that comes with seeing the chief findings as a part of considerably larger concerns.

Preparing the Survey Report. Recalling that the chief purpose of the survey is to justify status quo or to show the need of improvement (or both), the survey report assumes a position of importance. It serves the general purpose of the usual research report. It also serves as a public relations device or medium, if one of the secondary purposes of the survey is to influence at least some of the readers. Thus, each of the publics which is legitimately interested in the survey should be kept in mind when the report is prepared. This suggests, in turn, that each section of the report in its form, emphasis, and style might differ from the other sections.

Twelve survey-reporting techniques have been evaluated by four juries (8). Eleven of them received at least an 8.2 median rating from each jury, a rating of 10.0 being "high" (both desirable and practical). These highly rated 11 techniques follow: each phase of health and physical education covered in the survey should be written separately, with its own summary; a general summary of the entire survey should be made; explanation of unfamiliar terms should be included; sources and methods of collecting data should be stated; standard form in presenting tables, notes, and the like should be used; the status and needed improvement of the situation should be discussed; clarity, unity, and logic should be used in the organization of the report; the relationship and co-ordination between the survey data and their explanation should be established; the form and style should be made simple, direct, easy-to-read, and attractive; only one survey report should be made; data presented in graphic form should be explained; and charts, diagrams, and the like should be used to aid in ready understanding of findings.

In addition to these techniques, the writer of the report should keep in mind that the report bears the obligation of motivating and maintaining action by those who are to carry through the recommendations. Further, the amount of detail presented in the report depends chiefly on the exact purposes of the survey.

Such practical matters as size and shape of the report, type of paper, and cover are related to decisions made early in the planning stages, as are number of copies, distribution, and method of reproduction. Such decisions again are geared to the specific purposes of the survey and to available funds. Contracts for reproducing the report should be made with consideration for current depends chiefly on the exact purposes of the survey.

If the health, physical education, and recreation phases of a survey are only part of a larger survey, as they often are, such matters as have just been mentioned are seldom brought to the attention of the surveyor of the specialized fields. In any event, any member of the survey team should avoid becoming "an unnamed source" for press stories about the survey or any detail of it before the report is released or approval given by *all* proper authorities. The suggestion is implicit here that those responsible for the survey should fully and carefully plan press releases, public meetings if appropriate, and other suitable techniques of public relations in order that proper advantage be taken of the survey, the report of it, and its meaning. The carrying out of survey recommendations costs large sums of money. Those who pay these costs should be adequately informed, with timing as a special item of consideration.

Estimating Effectiveness. Evaluation of a survey's effectiveness is related to its purposes, although this relationship is not always recognized. Thus, some laymen and even some professionally

5

prepared personnel are unaware of the incongruity of attempting to estimate the effectiveness of a survey on the basis of other criteria. This is to say, then, that an early point made in this chapter—that purposes should be carefully formulated without the pressure of results-in-a-hurry—constitutes one of the most vital tasks of the surveyor. Its importance accounts for the continual emphasis in these pages on the keystone position that purposes have.

So well understood should be a survey's purposes, and so well supported by those interested in the survey, that any post-survey appraisal of the survey's effectiveness on any other basis would not and could not occur. It becomes implicit, then, that the representatives of many groups participate in the determination of these purposes.

Because each survey is unique and its purposes peculiar to itself, no formula for estimating its effectiveness can be furnished beforehand. Some secondary purposes of a survey may demand that testing and measuring of some operation be effected before that phase of the survey's effectiveness can be estimated. Some other purposes may demand the use of less precise tools and operations. Further, the need to select *appropriate* tools and methods of appraisal cannot be ignored, but it cannot be accurately predicted what these will be.

Neither the public nor the local personnel should be led to expect a solution of all problems. The possibility of such an outcome is sometimes unintentionally held out to the group in the effort to gain their co-operation and support.

The evaluation of the survey's effectiveness is in terms of its immediate outcomes and its long-range effects, but it must never be forgotten that these are anticipated in the survey's purposes. Again, it is clear that full and careful consideration should be given *at the outset* to the formulation of these purposes; and that the surveyor should avoid a possibly more attractive, more expansive, more spectacular set of purposes which cannot be met, with the resultant low rating of effectiveness.

Finally, the survey's effectiveness rests ultimately in the hands of local personnel. Therein lie the survey's strength and weakness, as far as the disposition of the findings and recommendations are concerned. Whether status quo is found to be justified or whether it is found that changes are needed, the responsibility for that disposition belongs to local persons. Herein, indirectly, is indicated one reason for the first point made in this chapter—namely, that the survey is subject to forces that may render it an unacceptable medium of *research.*

BIBLIOGRAPHY ON THE SURVEY

1. ALOIA, ALEX D. *The Organization of Student Recreation Programs in Selected Large Institutions of Higher Learning.* Unpublished Doctor's thesis. Los Angeles: University of Southern California, 1951.
2. BINGHAM, W. V., MOORE, B. V., and GUSTAD, J. W. *How To Interview.* Fourth edition, revised. New York: Harper & Brothers, 1959.
3. BLAND, HESTER. *An Analysis of the Activities of Indiana Public School Nurses Employed by Boards of Education.* Unpublished Doctor's thesis. Bloomington: Indiana University, 1956.
4. BROTEN, GEORGE A. *An Exploration of Swedish Physical Education.* Unpublished Doctor's thesis. Los Angeles: University of Southern California, 1957.
5. BURNETT, JOSEPH H., and O'BRIEN, FRED J. "Survey of Football Injuries in the High Schools of Massachusetts." *Research Quarterly* 4: 91-98; October 1933.
6. CASWELL, HOLLIS LELAND. "City School Surveys. An Interpretation and Appraisal." *Contributions to Education.* No. 358. New York: Bureau of Publications, Teachers College, Columbia University, 1929.
7. DANIELS, A. S. "Report on National Survey of Student Recreation in Colleges and Universities." *Research Quarterly* 11: 38-54; October 1940.
8. DAVIS, ELWOOD C. "Methods and Techniques in Surveying Health and Physical Education in City Schools." *Contributions to Education.* No. 515. New York: Bureau of Publications, Teachers College, Columbia University, 1931.
9. EVANSVILLE COMMUNITY CHEST AND COUNCILS. *Survey of Health Welfare and Recreation Services, Evansville, Indiana.* Evansville, Indiana: Community Chest and Councils, 1947.
10. FORNIA, DOROTHY L. *Coeducational Physical Education in Institutions of Higher Learning.* Unpublished Doctor's thesis. Los Angeles: University of Southern California, 1957.
11. FRYKLUND, VERNE C. *Trade and Job Analysis.* Milwaukee: Bruce Publishing Co., 1942.
12. JACKSON, C. O. "The Status of Physical Education in Accredited Schools of Illinois." *Research Quarterly* 9: 47; March 1938.
13. KERR, ROBERT. *The Status of Undergraduate Professional Preparation in Physical Education for Men in New England Colleges and Universities.* Unpublished Doctor's thesis. Bloomington: Indiana University, 1955.
14. KOOS, LEONARD V. *The Questionnaire in Education.* New York: The Macmillan Co., 1928.
15. LEE, MABEL. "A Survey of Athletic and Gymnastic Costumes Used by American Girls and Women." *Research Quarterly* 3: 5-47; March 1932.
16. NATIONAL EDUCATION ASSOCIATION, RESEARCH DIVISION. "The Questionnaire." *Research Bulletin* 8: 1-51; 1930.
17. NATIONAL EDUCATION ASSOCIATION, RESEARCH DIVISION. "Salaries and Salary Schedules of Urban School Employees, 1956-57." *Research Bulletin* 35:65-95; April 1957.
18. NATIONAL RECREATION ASSOCIATION. *Long Range Recreation Plans, Town of Kearny, New Jersey.* New York: the Association, 1942.
19. NATIONAL RECREATION ASSOCIATION. *Public Recreation, Gary, Indiana.* (Edited by L. H. Weir.) New York: the Association, 1946.

20. NATIONAL RECREATION ASSOCIATION. *Schedule for the Appraisal of Community Recreation*. New York: the Association, 1944.
21. PHILLIPS, MARJORIE. "Problems of Questionnaire Investigation." *Research Quarterly* 12: 528-37; October 1941.
22. SEHAM, MAX. "The Reliability and Validity of the Questionnaire Method." *Research Quarterly*, Supplement 5: 31-43; May 1934.
23. STEALY, RICHARD. *The Status of Driver Education in Selected Public Schools in Indiana*. Unpublished Doctor's thesis. Bloomington: Indiana University, 1957.

The Case Study

LAWRENCE RARICK

The case study is used to provide detailed information about an individual, institution, or situation. It is concerned primarily with determining the unique characteristics of the exceptional, rather than the attributes which are typical of many. The case approach has perhaps enjoyed most widespread use in medicine, law, and clinical psychology, for in each of these fields the practitioner deals with problems of a highly individualized nature. In the schools, the case method has been effectively used in the individual study and guidance of children with reading difficulties, speech problems, or psychological-emotional disturbances. However, few research studies using the case method have been reported in physical education literature, although it is a technique which coaches routinely employ in the critical analysis of the performance of their athletic teams.

Although the case study is most frequently used in the solution of individual problems, an accumulation of data from several similar cases frequently furnishes important data for comparative studies and for examining factors intimately associated with specific problems. For example, many advances in the field of medicine have come from careful study of case records of practicing physicians.

The case method has also been used effectively to study in detail the highly successful or unsuccessful person, as a means of identifying the traits which characterize him. While the presence or absence of the observed traits does not necessarily establish a causal relationship, identification of these characteristics does give the research worker something definite upon which to build. In the behavioral sciences where identification and isolation of basic factors in behavior problems is extremely difficult, the case method has been used effectively. In fact, in many fields of human inquiry where precise methods are not available for establishing cause and effect relationships, the case approach has provided sufficient evidence for establishing well-defined hypotheses concerning the interaction of associated variables. The case method, therefore, is an effective approach in resolving a particular difficulty, and frequently provides valuable data for formulating tentative generalizations concerning individuals or groups which are highly similar in some important respects.

The problem of delinquency is particularly well suited to the case approach, and as a result many case reports on delinquents are available. One of the early classical studies in delinquency is the report by Healy and Bronner (3) of a single case referred to the Judge Baker Foundation in Boston. The report includes a complete record of the personal, social, and environmental background of the child. A broader approach to the study of delinquency is illustrated by Harvey (2), who examined records of a large number of socially maladjusted American and Mexican boys in an attempt to gain insight into the physical, psychological, and social factors associated with delinquent youth. In a study more closely oriented to physical education, Sheldon (7) utilized the case approach in studying the physique of delinquent youth. On the basis of data on some 200 cases, Sheldon supports the belief that delinquency has definite biological roots.

Case studies in which the individuals were selected because of their unusual capacities or talents frequently furnish valuable information on the factors associated with these abilities. For example, Cureton (1) has provided a considerable body of individual data on 58 male athletes of national championship and Olympic caliber. This study gives information on the physical attributes, performance abilities, and organic efficiency of these men and provides some insight into the role these variables play in top-quality performance. Rarick and McKee (6) have presented case data on 20 children, 10 of whom were high achievers and 10 of whom were low achievers on a battery of motor tests. The findings provide information on the kinds of early experiences most closely associated with these children in the widely divergent groups.

The case approach is also an effective method for studying communities, schools, organizations, and the various institutions of our society. An excellent illustration of a comprehensive study of community life and the impact of social institutions upon the lives of adolescents is the work of Hollingshead (4). This study, providing data on some 735 adolescents growing up in a Midwestern community, points out the important role which family status in the social structure of the community plays in determining the social behavior of the adolescent in relation to the school, the church, recreation, his peers, and his family.

BASIC METHODOLOGY

The steps outlined below are usually followed in the conduct of a case study.

Determining Value. The investigator makes certain that the person, institution, or situation is sufficiently unique to warrant detailed investigation. If an investigation of this type is to be of real value, it should be directed toward the solution of a real difficulty or provide some insight into the organization of factors associated with some unusual phenomenon.

Obtaining Relevant Data. The investigator first obtains all data believed to be relevant to the problem. Where the problem pertains to a person or persons, the following sources of information may be used.

Health Examination. A complete medical examination may be necessary. However, the nature of the problem under investigation determines the type of health data which is needed. Frequently, special tests of vision, hearing, or nutritional status may provide valuable information on the problem.

Standardized Tests. Tests upon which norms have been developed are valuable in making estimates of the "normality" of the case under observation.

Personal Interviews. The interview provides a valuable source of personal data and is included as a basic research method in many case studies.

Observations of Behavior. Data of this kind may include recorded observations of play behavior, social behavior, or subjective judgment of physical performance.

Special Purpose Devices. It is often necessary to develop tests designed to measure particular traits or abilities. To investigate certain kinds of behavior, it may be desirable to record permanently certain aspects of the behavioral phenomenon. For example, Hubbard (5) used mechanical and electrical recording instruments in studying the difference between trained and untrained runners. Zimmerman (9) employed cinematographical analysis in studying the characteristics of skilled and unskilled performance in the standing broad jump.

Historical Data. Often information covering an earlier period of time is needed in order to interpret present status. This information may be obtained from permanent record files, documents (local or governmental), periodical sources, or personal interviews. Cumulative school records are a reasonably sound source of data for case studies. Care must be taken that all sources of an historical nature have been checked for accuracy. (See Chapter 14 on Historical Method.)

The investigator must also be sure that all data-collecting devices have been validated and checked for reliability. Finally, a critical review needs to be made of all data, both present and past, to ensure authenticity and accuracy of all relevant information.

Analyzing the Data. An understanding of the interaction of the variables at work in the situation under investigation can come about only after the data have been logically classified and subjected to appropriate methods of analysis. The logical organization and grouping of like data ordinarily present no problem. However, the limited number of cases and the nonrandomized character of the sample place restrictions on the statistical methods appropriate for these kinds of data. This does not reduce the effectiveness of the analysis, since the vast array of data on each case provides the opportunity to search for patterns of factors or events related to the phenomenon under investigation. In fact, the limited number of cases typical of the case study is well suited to a pattern analysis. This approach is particularly fruitful in gaining insight into the relative effect which different variables have on the present status of the case. Furthermore, intelligent use of data collected on the case at an earlier time may provide valuable clues in interpreting the findings. For example, Wetzel (8), in his discussion of growth failure in children, illustrates through case studies the need for interpreting present status in the light of the past. The analysis of case data utilizes all relevant information, past and present, which may help to explain the circumstances as they exist at the moment.

Making Recommendations. Frequently, case studies are conducted to throw light on a concrete problem or difficulty with the view to making recommendations for change or treatment. In such instances, the experiences gained in the successful treatment of identical or highly similar cases are useful in making recommendations for the future course of action. Care should be taken that an accurate record is kept of all procedures used in the treatment program.

On the other hand, the case approach may not involve treatment but may be used as a means of identifying the characteristics of persons who have demonstrated unusual ability in some line of endeavor. Obviously, in these instances, the concern is directed toward learning more about the clustering of traits under these conditions and to the possible interaction of variables which may have produced the desirable condition.

Appraising Effectiveness. The final step in the case study is the appraisal of the effectiveness of the recommended change. This may be accomplished by testing procedures, observational techniques, or special purpose devices of one kind or another. Ordinarily, the effectiveness of the instituted change cannot be adequately evaluated immediately after the program has been discontinued but must be appraised in terms of its more lasting effects.

VALUES AND LIMITATIONS

Caution must always be exercised in making generalizations about a single case. Statistically, an N of 1 provides little or no ground for scientific prediction, even though the case has been exhaustively studied. However, with many replications, the case approach becomes a powerful device for providing important information about personal and social phenomena which may be used to advance human knowledge in terms of basic understandings as well as of future courses of action.

BIBLIOGRAPHY ON THE CASE STUDY

1. Cureton, Thomas K., Jr. *Physical Fitness of Champion Athletes.* Urbana: University of Illinois Press, 1951. 458 p.
2. Harvey, Louise F. "The Delinquent Mexican Boy." *Journal of Educational Research* 42: 573-85; April 1949.
3. Healey, W. A., and Bronner, Augusta A., editors. *The Judge Baker Foundation Case Studies.* Series 1, Case Study No. 1. Boston: Judge Baker Guidance Center, 1922. 42 p.
4. Hollingshead, August B. *Elmtown's Youth.* New York: John Wiley & Sons, 1949. 480 p.
5. Hubbard, Alfred W. "An Experimental Analysis of Running and of Certain Fundamental Differences Between Trained and Untrained Runners." *Research Quarterly* 10: 28-38; October 1939.
6. Rarick, G. Lawrence, and McKee, Robert. "A Study of Twenty Third-Grade Children Exhibiting Extreme Levels of Achievement on Tests of Motor Proficiency." *Research Quarterly* 20: 142-52; May 1949.
7. Sheldon, William H.; Hart, Emil M.; and McDermott, Eugene. *Varieties of Delinquent Youth.* New York: Harper & Brothers, 1949. 899 p.
8. Wetzel, Norman C. *The Treatment of Growth Failure in Children.* Cleveland: Newspaper Enterprises Association Service, 1948. 102 p.
9. Zimmerman, Helen M. "Characteristic Likenesses and Differences Between Skilled and Non-Skilled Performance of Standing Broad Jump." *Research Quarterly* 27: 352-62; October 1956.

The Genetic Method

LAWRENCE RARICK

Status studies provide information on conditions as they exist at the moment. This is a satisfactory means of furnishing data for descriptive and comparative purposes and for examining interrelationships among variables as they are operating at that time. However, in many instances, the individuals or institutions under investigation are in the process of developmental change. This places restrictions on such inferences which can be drawn from a status study, particularly if the inferences pertain to the operation and interaction of variables over time. Under these conditions, the most suitable approach is the genetic method, in which observations on the variables in question are repeated at specific intervals over as long a time span as possible.

In the biological sciences, the genetic approach has been effectively used in studying the growth of plants and animals and, in conjunction with the experimental method, it has been an effective means of identifying and isolating specific growth factors. The genetic approach has also been employed in studying the physical growth and mental development of children, providing valuable information on the individual nature and variability of human development.

UNIQUE CONTRIBUTIONS

Until recently, most of the information on the growth and development of children was based on cross-sectional data accumulated on large numbers of individuals. Means and standard deviations for each growth variable by age and sex were thus available for plotting general growth trends. However, it soon became apparent that growth patterns of individual children were often substantially different from the growth curves plotted from normative data. Recognition of the marked individual differences in rates of growth and in the timing of developmental changes raised serious questions concerning the value and use of growth curves obtained from cross-sectional data. Therefore, research workers believe that it is more profitable to study intensively the growth and development of the same group of children over a period of several years.

The genetic method offers many advantages in studying developmental phenomena, particularly the growth and development of children. The plotting of individual growth curves provides a means of comparing rates of development of different growth variables for a particular individual, and also permits inter-individual comparisons. Furthermore, means and standard deviations can be computed at any point in time on all growth variables, as in the case of the normative survey. Determinations can also be made of growth increments for the individual and for groups. And finally, the interrelationships among specific growth factors which are operating over a time span can be determined, since all data are from the same children. Thus, the genetic method provides a much sounder basis for drawing inferences concerning the nature of development than do data collected on different children at different stages of development.

Perhaps the work of Terman (12, 3) in his *Genetic Studies of Genius* represents the most comprehensive genetic study of children yet published. In this study periodic observations were made on the growth of gifted children to ascertain those aspects of development which tended to differentiate gifted from normal children. Well over 1,000 children were included in the study, although the major portion of the report dealt with some 661 cases upon which more extensive data were collected. The information obtained on each child was extensive, including data on intelligence, health, education, racial background, recreational habits and interests, and hereditary and socio-economic background. The study was carried through a period of more than ten years, and a later follow-up investigation (13) provided valuable information on the status of the subjects in their social, business, and professional life as adult citizens.

The Harvard Growth Study (4) is an example of a longitudinal study of children, in which the growth of some 3,500 normal children was followed for as long as they remained in the elementary and secondary schools. This study was primarily directed to answering certain questions concerning the interrelationships among measures of physical growth, mental development, and maturational status as these children advanced toward maturity. The study clearly pointed out the individuality of growth careers and emphasized the need for caution in attempting to predict individual development from normative data.

In physical education and closely allied fields, only a limited number of genetic studies have been conducted. As a part of the California Adolescent Growth Study, Espenschade (5) and Jones (7) have provided valuable data on motor performance and strength development of a group of adolescent boys and girls. Espenschade included data on the developmental changes in motor performance of 80 girls and 85 boys as determined by six performance tests given at 6 months' intervals over a period of 3½ and 4 years respectively. Jones, using four measures of static dynamometric strength, observed patterns of strength development of 139 boys and girls over a 6½-year period. The work of Gesell (6) and Shirley (10) on the early motor behavior of infants demonstrated the value of the genetic method in observing sequential behavior patterns during early development. It should be noted that in the above studies care was taken that the measuring devices utilized were suitable for recording accurately the phenomena under observation at each point in the developmental sequence.

Only in this way was there assurance that valid interpretations of developmental change could be made.

One of the unique contributions of the genetic method lies in its potential for accurately predicting the development of individual children. For example, when growth curves are plotted for children of the same chronological age who differ markedly in their rates of sexual maturation, the trend for growth in body size and physical strength is distinctly different for the early and late maturers. Therefore, knowledge of the physical maturity status of the child, as well as information on other related factors, increases the power of predicting future developmental trends for individual children. This is borne out by the work of Bayley (1) in which growth curves, based on measurements of height and weight, have been established for a group of some 300 children on whom repeated measurements were taken from birth to 18 years of age. The grouping of data on children similar in physical maturity, from which growth curves were plotted, provided a much more accurate prediction of individual growth than could be obtained by growth curves established from cross-sectional data.

GENERAL METHODOLOGY

In the conduct of a genetic study, the following procedures are usually carried out.

Initial Planning. The investigator makes certain that the problem is suitable for employing the genetic method. This means that the hypotheses under examination can be most effectively explored by data secured on the same subjects or the same institution at regular intervals over a long period of time. In physical education, genetic studies offer a rich opportunity to study the role which such factors as physique, maturity, and strength play in the acquisition of motor skills as children advance toward maturity.

Careful planning is perhaps more important in genetic studies than in other kinds of investigations, since several years of continuous work are usually involved. This means careful planning not only in regard to the basic design of the study, but also in establishing the project schedule and in handling and processing the data. In the planning and conduct of the investigation, agreement should be reached on the exact points in time that the measurements are to be taken. Meredith (8) operates on a rigid time schedule of securing data within three days of each child's birthday. There should be assurance that all measures are valid and reliable and that all testers are trained to ensure objectivity and accuracy of measurement. Extreme care must also be taken in standardization of all measurement procedures.

Collection and Recording of Data. The data-collecting devices for genetic studies are similar to those described in the section above on the case study. However, great care needs to be taken in the selection of tests and measuring devices, since the recording of developmental change requires highly accurate tools of measurement. This is particularly important when the time intervals between measurements are short and the amount of growth small.

The data collected in a longitudinal investigation assume large proportions, and hence a systematic method is needed for recording and filing all data. It is recommended that a folder be kept on each child and that all test scores be recorded on permanent blanks and immediately filed in the child's folder. It is particularly important in the genetic study that the exact date of every observation be appropriately recorded.

Periodically, outside checks may be made on the characteristics of the study group by running controlled observations on other samples of children. This may tell the investigator something of the influence which drop-outs may have had on the characteristics of the study group.

Treatment of Longitudinal Data. One of the major difficulties in genetic studies is the problem of processing and treating serial records. The plotting of individual growth curves for specific growth variables is one approach that has been used in individualizing the treatment of developmental data. Obviously, this approach has its limitations, especially if one attempts to plot on a single graph data for several subjects recorded in different units.

Several methods are available for transforming raw data expressed in different units into equivalent values so that the data can be readily plotted. Olson (9), in his longitudinal study of children, converted each child's scores on each variable into an age unit based on norms for each variable. The use of this technique is open to question since there is little likelihood that the age variables have the same standard deviations, and hence the computed "ages" are not necessarily comparable. Conversion of raw scores into standard scores provides a more satisfactory means of establishing each child's relative position in the group for each growth variable. This is not only effective for plotting individual and group curves, but is also useful in developing a profile of traits for individuals at various points in development. Standard scores have been used effectively in handling growth data by Jones (7), Bayley (2), and Sontag (11).

A method of treating data which gives insight into the relative velocity and timing of growth is to convert raw scores into percentages of terminal status. Jones (7) used this approach in examining sex differences in growth of different strength variables during adolescence. Patterns of growth of different variables for both individuals and groups are frequently analyzed in terms of increments based on percentage rate of growth. The problem of analyzing longitudinal data can in part be resolved by the use of correlation methods, so that the relationship among variables over points of time can be determined, as well as interrelationships among variables at specific points in the growth cycle.

Interpretation of Findings. Although longitudinal data do provide valuable information on changes occurring in an individual over well-defined time intervals, the factors causing these changes may not be easy to identify. Caution must be exercised in drawing unwarranted conclusions, for the association of variables over time does not necessarily establish a causal relationship.

In conducting longitudinal studies, care must also be taken that inferences are drawn only for the age levels fully encompassed by the study. As Meredith (8) has pointed out, generalizations have been made on growth which have been based on extrapolation backward as well as forward, rather than within the boundaries of the study. Likewise, there is danger that the age limits set up by the study may result in a segmental approach to the study of development and in the use of truncated trends in the analysis of individual or group trends; that is, curves classified as linear might well have exhibited an acceleration phase, had the study been conducted over a longer time span. As in other methods of research, inferences drawn from genetic data should be confined to the sample studied or to the population from which the sample was drawn.

DIFFICULTIES

The extended period of time required to conduct a longitudinal study tends to discourage many research workers. This is particularly true of the graduate student who finds the pressures of time and finances a real problem in graduate study. However, in institutions where longitudinal growth studies are underway, students have made valuable contributions through their affiliations with on-going projects in studying the relationships and interactions among growth variables at designated points in the developmental phenomenon.

The maintenance of an intact group of subjects for the duration of the project is a major problem. By selecting subjects who are likely to be permanently located, the researcher can reduce this problem, but he also increases the likelihood of establishing a biased sample. In spite of the risk of drop-outs, it is usually better to use an unbiased sample from which inferences can be legitimately drawn. Co-operation of parents, school personnel, and the children themselves must be enlisted and maintained, if the project is to prove successful. Where the same performance tests are repeated periodically, care needs to be taken that identical conditions and testing procedures prevail. At times, differential factors of motivation, hour of testing, or season of testing may distort the data.

NEED IN PHYSICAL EDUCATION

Many important questions in physical education will remain unanswered until carefully conducted studies have been made on the same group of individuals over long periods of time. The long-term effects of exercise upon humans have not yet been adequately explored, although there is a tendency to draw inferences from mortality tables based largely on normative data. Little is known concerning the age and maturity levels at which skills can be most economically learned. Answers to these important questions will depend upon combined experimental-genetic studies. Although the genetic approach presents many difficulties and problems, it perhaps offers the most fruitful approach for obtaining answers to some of the most significant questions which confront physical education today.

BIBLIOGRAPHY ON THE GENETIC METHOD

1. BAYLEY, NANCY. "Growth Curves of Height and Weight by Age for Boys and Girls, Scaled According to Physical Maturity." *Journal of Pediatrics* 48: 187-94; February 1956.
2. BAYLEY, NANCY. "Individual Patterns of Development." *Child Development* 27: 45-74; March 1956.
3. BURKS, BARBARA S.; JENSEN, DOROTHY W.; and TERMAN, L. M. *Genetic Studies of Genius: III. The Promise of Youth*. Stanford, Calif.: Stanford University Press, 1930. 508 p.
4. DEARBORN, WALTER F., and ROTHNEY, JOHN W. M. *Predicting the Child's Development*. Cambridge, Mass.: Sci-Art Publishers, 1941. 360 p.
5. ESPENSCHADE, ANNA. *Motor Performance in Adolescence*. Monographs of the Society for Research in Child Development, Vol. 5, No. 1. Washington, D. C.: the Society, 1940. 126 p.
6. GESELL, ARNOLD, and AMES, L. B. "The Ontogenetic Organization of Prone Behavior in Human Infancy." *Journal of Genetic Psychology* 56: 247-63; June 1940.
7. JONES, HAROLD E. *Motor Performance and Growth*. Berkeley: University of California Press, 1949. 181 p.
8. MEREDITH, HOWARD V. "Longitudinal Anthropometric Data in the Study of Individual Growth." *Annals of the New York Academy of Sciences* 63: 510-27; November 1955.
9. OLSON, WILLARD C., and HUGHES, B. O. "The Concept of Organismic Age." *Journal of Educational Research* 35: 525-27; March 1942.
10. SHIRLEY, MARY M. *The First Two Years; A Study of Twenty-Five Babies*. Vol. 1, *Postural and Locomotor Development*. Minneapolis: University of Minnesota Press, 1931.
11. SONTAG, L. W., and REYNOLDS, E. L. "The Fels Composite Sheet, I." A Practical Method for Analyzing Growth Progress. *Journal of Pediatrics* 26: 327-35; April 1945.
12. TERMAN, L. M., and OTHERS. *Genetic Studies of Genius: I. Mental and Physical Traits of a Thousand Gifted Children*. Stanford, Calif.: Stanford University Press, 1925. 648 p.
13. TERMAN, LOUIS M., and ODEN, MELITA H. *Genetic Studies of Genius: IV. Twenty-Five Years' Follow-Up of a Superior Group*. Stanford, Calif.: Stanford University Press, 1947. 448 p.

CHAPTER 10

The Experimental Method

MARJORIE PHILLIPS
LEONARD LARSON

THE EXPERIMENTAL METHOD IS APPLIED TO PROBLEM SOLVING when factors needed for the solution of the problem may be controlled and their influence determined. Controls may be applied in a *laboratory setting* or in an *operational setting*. In the former, variables are held constant by precise experimental procedures in order to determine the influence of a single variable or combinations of variables. In the latter, the variables are allowed to operate in a natural and normal setting. This procedure is frequently more effective in the solution of educational problems. In both instances, certain basic controls—such as the identification of the population, the selection of the subjects, establishment of the duration of the experiment, and the elimination of extraneous influences—must be established before experimentation may begin. Finally, measurements are secured and effects are determined through logical and quantitative analysis.

The experimental method may be defined as a method of research designed to determine influences, both qualitatively and quantitatively, on a given phenomenon, or to determine influences between or among variables. It is the only method of research whose design demands controlled observation. The conditions under which the phenomenon is to be studied become the starting point for research.

PLANNING THE EXPERIMENT

The Problem. Probably the most difficult phase in the planning of an experiment is the identification and definition of the problem. Judgment determines the nature of the problem itself; what variables will be included or excluded; the place where the experiment will be conducted; the conditions under which the experiment will be administered; the time and duration of the experiment; the subjects to be used to represent a given population; the degree of precision needed to yield results worthy of generalization; the instruments to be used to measure the influnces of the variables; the nature and extent of pilot investigations before beginning the experiment; the pattern of control—the single group, the parallel group, or multiple groups; the method of equalization—by individuals or by groups; and the analysis design which most effectively demonstrates the results of the experiment.

In addition to the proper identification of the problem, the preliminary planning must also include an analysis and further identification of the problem by statements of hypotheses. Such statements give more precisely what is to be studied; the experiment is more sharply defined. Statements of hypotheses begin with what is assumed to be true, based upon avail ''e evidence. Research will either confirm or reject the hypotheses. The procedure is helpful in reducing the scope of the problem to a particular condition or circumstance.

It is also a part of the planning to determine what assumptions underlie the experimental plans. False assumptions yield invalid results. It is in some instances desirable to investigate the assumptions before beginning the experiment. For example, assuming that a test will measure an ability, when reliable evidence to support such an assumption is lacking, is indeed a questionable procedure.

Other steps in the process of defining the problem include plans for the collection of data and the selection of instruments. The heart of the experiment lies in the instruments. These must fit the group and yield data needed for the solution of the problem. The results will be valid to the degree that the instruments are valid. Procedures for the qualitative or quantitative analysis of data must also be planned before experimentation begins. The refinement of instruments may be necessary after a review or analysis of the procedures available or desirable.

Verification is desirable in all research, but it is particularly so in experimental research. This is due to chance conditions which may exist. Similar results with replications of the experiment will provide additional support for the conclusions.

Delimitations. One may attempt to eliminate from the experiment those variables which are not necessary for the solution of the problem. One may, for example, delimit an experiment to the male sex only. This procedure is desirable when such variables do not contribute to the solution of the problem but simply add to its complexity. Delimitations in educational research may include such factors or conditions as sex, age, institutional levels, type of agency or institution, quality of programs, religion, race, physical condition levels, and experience.

Subjects. The results of experimentation should lead to generalizations within the limitations of the experimental design. All individuals are different in their traits, characteristics, abilities, and reactions. The primary concern of the investigator in selecting his subjects is that they may be truly representative of the population about which he wishes to generalize. The techniques used to select the sample must be carefully described and must be such that the possibility of bias is eliminated. The subjects as well as the investigator must also be without bias. The subjects must be appropriate to the solution of the problem. For example, men majors in physical education are not representative of college men in general, or adult women as subjects would not be appropriate if the results are to be applied to children.

Selection of the subjects should begin with a definition of the population—as, for example, first grade boys. This is followed by definition of the characteristics of this population which are relevant to other pertinent factors in the experiment.

Other considerations of importance are availability of the subjects and the number of subjects. It is wise to have some assurance that subjects will be available throughout the duration of the investigation, and it is far better to have a few subjects with careful controls than large numbers with lax controls.

Experimental Design. Experimental research may be conducted on an "individual by individual" basis or on the "group by group" basis. If results are going to be applied to the individual, the former is the correct design. If, however, the results are to be applied to a population or group, as a whole, the latter is the acceptable pattern. The problem is one of definition and equation.

The basic patterns in experimental research, aside from starting with an individual, are the single group, the equivalent groups, and the rotational groups. The single group design is valid if the results are used to describe or characterize the group. If several variables are applied to the single group, for the purpose of comparison, the conditions at each application must be identical and the group must not be conditioned by the preceding variables. The equivalent pattern consists of two or more groups made comparable by the selection of subjects or by statistical treatment. The equating factors are only those which influence the experiment—for example, training in a physical fitness experiment. Generally, one group serves as a control group while the other serves as the experimental group. The rotational group pattern is a combination of the two; the application of the experimental variable is rotated between and among groups. The design is valid when changes produced by the experimental variable are not "carried-over" from one group to the other, or, if they are carried over, are neutralized through the process of rotation.

PRINCIPLES OF EXPERIMENTATION

Statistical Hypotheses. The basic thinking which leads to statistical inference is well expressed in the words of Fisher:

> . . . I have assumed, as the experimenter always does assume, that it is possible to draw valid inferences from the results of experimentation; that it is possible to argue from consequences to causes, from observations to hypotheses; as a statistician would say, from a sample to the population from which the sample was drawn, or, as a logician might put it, from the particular to the general. (11:3)

The foundation of modern experimentation rests on the skill of investigators in formulating fruitful hypotheses and designing experiments to test them. If the guidance provided by a productive hypothesis is lacking, the experimenter is foredoomed to an unsuccessful investigation.

The development of the hypothesis demands an understanding of the principles of the scientific method and a complete understanding of the problem being investigated. The hypothesis must be clearly and carefully formulated at the primary stage of the research and in such a manner as to permit the solution of the practical problem involved. The theories and ideas gained through the researcher's own experience, and that of others, provide the basis for the hypothesis. Once the hypothesis has been formulated, the subsequent plans for the investigation are developed in such a way as to render possible a verification of the hypothesis through direct observation. Verification of the hypothesis means that satisfactory evidence is provided to indicate its reasonableness or unreasonableness. The concept of proving a hypothesis is fallacious; rather, if the idea of proof is to be suggested at all, it should be thought of as an attempt to disprove the hypothesis. The hypothesis gives direction to the nature of the data to be collected and the manner in which they should be collected; it determines the techniques for organizing and analyzing the data; and finally it suggests the judgments to be made and the conclusions to be drawn. Any failure to pose the proper hypothesis relative to the problem under investigation may well result in a serious lessening of the efficiency of the investigation or cause the collection of data from which no valid inferences may be drawn.

In essence, statistical hypotheses are concerned with the probability that certain observed effects have resulted from chance factors (errors in random sampling) rather than from some differential treatment associated with the investigation. In stating his hypothesis, the investigator follows the law of parsimony and selects the simplest possible hypothesis, i.e., that there is no difference in the effect of the treatments under observation. This is known as the *null hypothesis* and it may be stated in other ways—for example, the difference between means of the treatment groups is zero, or the groups are random samples from the same population.

After a clear, unambiguous statement of the hypothesis has been developed, the subsequent planning and conduct of the experiment must lead to the point where it is possible to apply a test of significance to the results. The test of significance permits two possible outcomes for the investigation.

1. *Accepted Hypothesis.* In the instance where the hypothesis is accepted, the investigator subscribes to the belief that the ob-

served effects could reasonably be attributed to chance. The available evidence indicates that the hypothetical condition could be the true condition. The critical point at which the investigator no longer considers the hypothesis tenable must remain a matter of option, dictated by the exactness demanded by the investigator for any particular experiment. Obviously the experiment must fail in its purpose if the investigator cannot be satisfied with any specified probability. The standard accepted by many experimenters is the 5 percent probability level, although on occasion a much higher level is demanded.

2. *Rejected Hypothesis.* In the instance where the hypothesis is rejected, the investigator considers the observed and hypothetical effects to be incompatible. The probability of the hypothetical effect existing under the observed conditions is too small for reasonable expectation. If, for example, the confidence level used in rejecting the hypothesis were 5 percent, it would mean that the hypothetical probability, as determined from the null hypothesis under test, is only 5 in 100. It provides an expression of the investigator's credence in the hypothesis. The rejection of the hypothesis does not mean that this decision cannot possibly be reversed by future evidence. It simply indicates that from the evidence available from this particular experiment the hypothesis is unacceptable, and that to believe in this hypothesis would require adherence to a belief that an exceptional event had occurred.

Significance Levels and Classes of Error. When experimental results are evaluated through a test of significance, there is always the possibility of the conclusions being in error in either one of two directions. Those errors occur in the cases where a true hypothesis is rejected or a false hypothesis is accepted.

When a hypothesis is rejected, the exact probability of error is specified—and thus the experimenter makes known the risk involved in assuming that a significant result has been demonstrated. When a hypothesis is accepted, however, no declaration of the risk involved can be made since the relative frequency of error in retaining a hypothesis depends in practice on the discrepancy between the hypothetical true condition and the actual true condition. The closer the hypothesis is to the truth (without the two actually being coincidental), the greater the risk of accepting a false hypothesis; while the greater the disparity is between the hypothesis and the truth, the less the risk of accepting a false hypothesis.

Both types of errors should be considered when selecting a level of confidence. If a very high level of confidence is established for an experiment, the chance of accepting a false hypothesis is increased; and conversely, as the level of confidence is lowered, the chance of rejecting a true hypothesis is increased. Many investigators have been more concerned about rejecting a true hypothesis than accepting a false one, and so have insisted on a very high level of confidence. Actually, the decision should be guided by the probable consequences of making an error in one direction or the other. There are situations where it may be preferable to reject a true hypothesis and conclude there is a difference when none exists, rather than to accept a false hypothesis and conclude there is no difference when actually a difference does exist.

For example, in teaching a motor skill, a teacher may well wonder if the use of kinesthetic cues will improve achievement. A controlled investigation involving two random groups reveals that on the basis of average achievement the group receiving the kinesthetic cues scored several points higher on the criterion measure of the skill. Under the null hypothesis it is found that sampling errors of the size obtained would occur 5 times in 100 by chance alone. If the investigator insists on the 1 percent level of confidence, he may very well be accepting a false hypothesis and thus decide to eliminate the use of kinesthetic cues. On the other hand, suppose he rejects the null hypothesis at the 5 percent level of confidence and assumes the greater risk of rejecting a true hypothesis. This increased risk is probably well worth taking, since the use of kinesthetic cues involves no additional expense, in any application of the term, and could make some considerable contribution to achievement. If there is a health risk or if major expenditures of time, space, effort, or financial outlay are involved, the investigator may well insist on the higher level of confidence rather than take the chance of making a recommendation which, if in error, could be costly.

In general then, each situation must be examined before determining the best level of confidence to use. The primary concern in rejecting a hypothesis when it is true is that further experimentation may be based on that result in the belief that some condition has been established. The major danger in accepting a hypothesis that is false is that a certain line of experimentation which would produce fruitful results may be abandoned.

Randomization. The practice of using random groups in experiments is one of the fine advancements in modern research methods. In the words of Cochran and Cox,

> Randomization is one of the few characteristics of modern experimental design that appears to be really modern. One can find experiments made 100 or 150 years ago that embody the principles that are now regarded as sound, with the conspicuous exception of randomization (7:7).

The primary purpose of randomization is to ensure that no bias, known or unknown, will occur because of the manner in which the groups are selected. The selection of a sample in such a manner that it fails to meet the criterion of randomness imposes a limitation on the interpretations which may be derived from it. The validity of statistical inference is dependent on the fact that the sample is a random one, since otherwise no dependable estimate of the population parameter may be developed from the sample. (See Chapter 4.)

In evaluating the results of an investigation, one of the commonest practical questions which must be answered is whether an observed difference in the means of two groups is the result of the experimental treatments or whether it may be attributed to a difference in the initial ability of the two groups of subjects to respond to the treatments. If the subjects have been assigned to the groups at random or have been drawn at random from a common population, this question may be readily answered. The sampling errors arising from initial differences in subjects selected at random are measurable. It is, therefore, possible to determine the probability that an observed difference between two groups is due to chance factors. If this probability is high, then the differential treatments would not be given considerations as a cause of the difference. If the probability is low, it is then reasonable to proceed with an examination of the treatments as a possible cause of the difference.

When some nonrandom technique is used for assigning the subjects to the groups, a bias of the investigator—either conscious or unconscious—or some biasing factor in the method used for composing the groups, may be operating so that the groups are quite different in their initial abilities to respond to the treatments.

The sampling error arising from such biases is not measurable; hence, the interpretation of the results is obscured. Even in cases where there appears to be no possibility of a bias affecting the results, it is well to use the safeguard of randomization and thus provide an insurance against any unforeseen circumstance which might seriously disturb the results of the investigation.

Random selection does not necessarily imply the simple selection of a random sample from an identifiable population. In fact, it is quite rare that an investigator whose subjects are human beings can draw his samples in this manner. The experiment can be planned, however, to meet the assumption of randomness. The most commonly used techniques are to draw the subjects at random from those available or, where numbers are small, to use random techniques to assign all available subjects to the treatment groups. Any strictly chance method may serve for this purpose, but one which is particularly recommended is the use of tables of random numbers (13:114). In such cases as these, where it has been impossible for the investigator to draw samples at random from the total population with which he is concerned, it becomes necessary to define a hypothetical population. The hypothetical population is described as one in which all the members are like those involved in the investigation. All interpretations from such an experiment are relevant only to this hypothetical population. If generalizations to other populations are attempted, they will be primarily a matter of speculation tempered by judgment and experience, since there will be no statistical safeguards to support the inferences.

Investigators sometimes design experiments in such a way that no meaningful interpretations can be made. Suppose the efficiency of two different types of grip strength testing devices is to be compared. A group of subjects is available, all of whom are tested first on device number 1 and then on device number 2. The average grip strength registered is found to be higher when the subjects are tested by device number 2, and the difference is statistically significant. The investigator now finds himself in a quandry, since he has no way of knowing whether the difference may be attributed to a difference in the effectiveness of the devices or to such factors as greater familiarity with the testing situation and increase in strength resulting from the practice obtained on device number 1. This problem could have been resolved, had the investigator designed his study to incorporate random techniques. The obvious source of bias, and other unsuspected sources of bias, could have been eliminated by the simple expedient of flipping a coin as each subject appeared to be measured and letting the fall of the coin decide whether the subject would be tested first by device number 1 or device number 2. While the problems in random sampling are not always solved this simply, the situation serves to illustrate the protection afforded by random techniques in the elimination of bias.

Basic Sources of Error. If the effect of possible sources of error on experimental results is to be assessed, a very careful preplanning of every stage of the investigation is demanded, as well as clear insight into the probable factors which could influence the results.

While it is obvious that gross errors in measurement and calculations are possible, these sources of error will not be discussed here, since the experiment is not designed to study such errors and it is unthinkable that any serious investigator would permit them to occur.

Random errors of measurement will also be given little consideration, since they will tend to cancel out. It is only when errors of measurement are systematic that they may seriously affect the results, and again, the good investigator will make every effort to reduce these errors to a minimum.

In any experiment involving an evaluation of an observed difference, the primary consideration is whether the difference is a real one, and if so, that this difference may be attributed to the treatment effects. Before the conclusion may be reached that a real difference probably exists, it is necessary to show that the difference is not reasonably attributable to sampling error arising out of the random selection of the subjects.

Intersubject Errors. The first basic source of error will be referred to as "intersubject errors." A valid estimate may be secured for intersubject errors in any experiment where random techniques have been used in drawing the samples. No valid estimate of intersubject errors is available if any method other than a random one is used in selecting sample members. Thus, in the properly designed investigation, when a difference is being analyzed, it will be possible to state the probability that a sampling error of the size obtained can be attributed to chance. If the sampling error is shown to be too large to be attributed entirely to chance intersubject differences, and hence resulting from something other than the random selection of the subjects, there remains another possible source of error to be considered.

Intergroup Errors. This second basic source of error will be referred to as "intergroup errors." Intergroup errors occur when some factor (or factors) is allowed to operate so that it affects all the members of one experimental group in a consistently different way than it affects the members of the other experimental group or groups. The effect of this is a difference between experimental groups which is in no way connected with the differential treatments. Let us suppose that an investigator is concerned with the problem of comparing the effectiveness of two methods of teaching swimming to children. There are certain factors which quite obviously could be responsible for introducing intergroup errors unless they were controlled. The amount of practice allowed on the various skills, the duration and number of the practice periods, the time of day, the quality of the instruction, the general environment, and facilities available are all readily identified as possible sources of intergroup error, unless all groups are treated alike in these respects. These same kinds of factors would doubtless be considerations in any learning experiment, but the investigator must also be alert to factors which are unique to a given problem. Variations in water temperature, for example, could definitely affect the results in the swimming experiment. Assuming that some specific temperature would be best for optimum results, any variation from that for one group might affect the learning. This could be particularly important at the time criterion measures are being secured. If one of the groups becomes chilled because the water is too cold, that group may have considerably lower scores on the criterion measure when actually no real difference exists between the average swimming skill of this group and the other groups of the experiment.

For single replications of an experiment, there is no error estimate available to measure intergroup errors; hence, it becomes the responsibility of the investigator to identify these factors during the *planning* stages of the experiment and make provision for keeping them constant across the groups.

The treatment effects may, then, be considered to be the cause of a difference in means of experimental groups when (a) a valid estimate of intersubject errors is available and the difference is shown to be a statistically significant one, and (b) all sources of intergroup error have been identified and kept the same for all groups.

Precision. When subjects are assigned at random to groups, it is possible that just by chance the subjects of the various groups may be quite different from each other in their capacities to respond to various experimental treatments. When these differences in initial abilities are related to performance on the criterion measure, the effect is to inflate the estimate of errors beyond what it would be if the initial abilities of the subjects of the groups had been more alike.

In order to increase the precision of the error estimate, it is frequently possible to select a design which will reduce the variation of subjects because of chance differences in initial ability. This purpose may be accomplished by matching the subjects of the groups, prior to the administration of the treatments, on the criterion trait or some trait related to the criterion trait. The higher this relationship, the more effective will the matching be in increasing precision. Once the matching process has been utilized, it is of utmost importance that the proper estimate of errors be applied, i.e., one that will not give consideration to the source of error that has been eliminated. It would be better not to attempt to make the groups more alike and to use a less precise error estimate, than to make the groups more alike and fail to use the appropriate estimate. A simple illustration of this is in an investigation in which the subjects of two groups are paired in terms of their initial ability on the trait being studied. The difference between the means of the two groups on the final criterion measure is then incorrectly evaluated by the estimate of errors for independent random samples. The incorrect error estimate will not reflect the increased precision gained through the matching process, but the difference between the groups will be reduced. This experiment may thus appear less conclusive than it would have if no matching had been done.

A second design for introducing increased precision is one which permits the results to be analyzed by the methods of covariance. By these methods the variations of the subjects in initial ability are equalized statistically.

The size of the samples used in an experiment will also affect the precision of the error estimate. Other things being equal, the larger the sample is, the higher the precision. It should be obvious, however, that there are many other considerations that would govern sample size. An increase in sample size which becomes excessive in terms of the cost of the experiment would hardly be considered worthwhile. In addition, there are many cases in which an increase in sample size beyond a certain point would so affect the efficiency of the investigation as to render the results practically worthless. For example, when the extent of learning of some motor skill is the criterion measure, the increase of the groups beyond a certain size could so affect the teaching-learning situation as to result in no gain in skill regardless of the experimental treatments applied. The standard for sample size must in the final analysis be governed by the purposes and requirements of a particular investigation.

Assumptions. If clear and valid statistical interpretations are to be made of the results of an experiment, it is essential that the assumptions underlying error estimates be met. There is no defensible reason for accepting such assumptions as true without any precautionary measures or without any supporting evidence of their truth.

In general, assumptions may be divided into two categories according to the techniques used for demonstrating that they have been satisfied.

The first category would include all assumptions that can be met in the design of the investigation. For example, one of the commonest of all the assumptions is that the samples are drawn at random from a common population. It is possible for the investigator to satisfy this assumption by using random techniques in the selection of his subjects. A second example of this class of assumptions occurs in the analysis of covariance. It is assumed that the initial measures are unaffected by the experimental treatments. The truth of this assumption may be easily assured by the investigator. It simply demands that he secure all initial measures preliminary to the treatment period.

The second category would include assumptions over which the investigator has no control in the planning of the investigation, but which may be evaluated by a test of significance. This category is illustrated by such assumptions as homogeneity of variance and linearity of regression. The minimum requirement would be for the investigator to report the probability resulting from the test of significance, and if possible, he should attempt to assess the probable effect on the results of a failure to satisfy an assumption.

SIMPLE EXPERIMENTAL DESIGNS

If an experiment is to be considered a good one, there are many essential characteristics which it should possess. These are so numerous that it would be beyond the scope of this book to attempt to identify them all. However, Lindquist has developed a summary which describes the most important characteristics of a good experiment, and which are given as follows:

1. It will insure that the observed treatment effects are *unbiased* estimates of the true effects.
2. It will permit a quantitative description of the precision of the observed treatment effects regarded as estimates of the "true" effects.
3. It will insure that the observed treatment effects will have whatever degree of precision is required by the broader purposes of the experiment.
4. It will make possible an objective test of a specific hypothesis concerning the true effects; that is, it will permit the computation of the relative frequency with which the observed discrepancy between observation and hypothesis would be exceeded if the hypothesis were true.
5. It will be *efficient;* that is, it will satisfy these requirements at the minimum "cost" broadly conceived. (22:6)

There are many experimental designs applicable to the solution of a variety of problems. Some of the simpler designs, all concerned with single replications of the experiment, will be discussed and illustrated here. The discussion will be limited to the type of educational experiment which examines the effect of differential treatments on the arithmetic average or mean of a criterion variable.

Single Group Design. The simplest possible design would be one in which a single group was studied over a period of time in an attempt to observe the effect of some treatment. In this case, an initial test of the criterion variable would be administered to a random sample of some population, either real or hypothetical.

The treatment would be applied over a reasonable period of time and a final test of the criterion variable would be administered. The test of significance would be applied to the difference between means of the group on the initial and final tests, giving consideration in the error estimate to the relationship which doubtless would exist between the initial and final test measures. Both the advantages and limitations of this design seem quite evident. Many mechanical difficulties are obviated by using a single group, and intersubject errors are eliminated, but since only one group is involved, there is no basis for comparison in the logical analysis of the results. The error estimate is valid and precise, but the cause of a statistically significant difference may be obscure. The investigator may believe that the treatment is the cause, but he may also experience difficulty in eliminating other factors which could reasonably have been the cause.

Let us consider a case in which an experimenter is interested in learning whether significant gains in the endurance of children are made following a certain type of training program. An endurance test is given to the subjects prior to the training period and following the training period. A significant difference is found between the initial and final mean endurance scores, with the final mean score being the higher of the two. It may be validly concluded that following the training period the group is superior in endurance to what it was prior to the training period. However, the reason for this gain cannot be clearly demonstrated. It may be due partially or entirely to the training, or it may be due to natural maturation gains, to the practice effect of repeating the test, or to physical activities not controlled.

The single group technique may have useful applications in situations where two or more variations of a variable are to be studied, the variations of the variables constituting the treatments. The various treatments are administered in sequence to the subjects, and as a result errors arising from intersubject differences are eliminated. This approach is limited to situations in which the effect of one treatment does not influence subsequent treatments.

Let us suppose that a swimming coach is interested in the effect of various dosages of oxygen on the performance of champion short distance swimmers. A group of champion swimmers is selected at random from a defined population of champion swimmers. Each of the selected subjects swims the specified distance twice—once preceded by one dosage (A) of oxygen and once preceded by another dosage (B) of oxygen. If all subjects are at their performance peaks, it may be reasoned that no practice effect will be present and hence all subjects can receive dosage A on the first swim, and dosage B on the second swim. In order to eliminate any possible bias from this source, the administration order of the dosages for each subject can be determined by a chance method. An analysis of the mean difference of the two performances of the subjects may be made and validly interpreted. The main advantage of this design is the increased precision resulting from the elimination of intersubject errors. However, since the treatment effects in the majority of cases are not independent of each other, the usefulness of this design is limited to the exceptional situation.

Simple Random Designs. These may consist of two independent groups or of three or more independent groups, as described below.

Two Independent Groups. This design is useful for observing the effect of a certain treatment on a group, with a second group acting as a control and receiving no special treatment, or for observing the comparative effect of two treatments. The treatments would be applied over a specified period of time, and finally a criterion test would be administered.

To illustrate this design, let us suppose that an investigator wishes to examine the relative effect of two bouts of exercise on the physical condition of young men. The subjects available for the experiment are assigned at random to two groups, and the groups are assigned at random to the two bouts of exercise. At the end of the treatment period, a physical condition test is given to each of the subjects, and the difference in the mean physical condition of the two groups of young men is evaluated. The error estimate secured from the statistical analysis provides a perfectly valid estimate of the intersubject errors, but it may lack essential precision. If a significant difference is demonstrated, all is well. However, it is possible that just by chance the subjects of the two groups were quite unlike each other initially, thus expanding the error estimate to the point that a difference actually existing is not revealed as significant. In situations where a more precise design can be used, this design would not be preferred.

Three or More Independent Groups. When several variations of the same variable are to be observed in the same experiment, subjects may be randomly assigned to the several groups and the groups, to the various treatments. The same general procedures would be followed as for two independent groups. The statistical analysis would be by the methods of analysis of variance.

This may be illustrated by the above study on the bouts of exercise with the exception that, instead of two bouts, several bouts would be selected for use as the experimental treatments. The error estimate derived from the analysis of variance would be perfectly valid for evaluating intersubject errors but again might lack sufficient precision to reveal differences which actually exist.

Related Group Designs. Since the simple random designs may lack desirable precision, attempts may be made to increase precision by the use of techniques which reduce initial dissimilarities in subjects from group to group.

Two Paired Groups. In this design the subjects of the investigation are paired, person for person, on the basis of an initial measure which is either the same as, or is highly related to, the criterion measure. After the pairing is completed, the members of each pair are assigned at random to the groups and the groups, to the treatments. The administration of the treatments is then followed by a test of the criterion measure.

To illustrate this design, let us consider a situation in which the investigator wishes to study the effect of two different diets on increasing the weight of undernourished children. Let us assume that the investigator is in a summer camp for such children and hence may control quite carefully all of the living conditions. The initial weights of the children will obviousl[illegible] an important factor in the evaluation of their final weights, so it is important that the two groups be made as nearly alike as possible in terms of initial weight. Each child is weighed preliminary to the treatment period; then on the basis of weight the children are paired as closely as possible. The two members of each pair are assigned to the two groups at random, and the two groups are then assigned to the two diets at random. During a specified time the children follow the experimental diets, and at the end of this time they are weighed once more. The mean of the differences in final weight of the pairs is evaluated, and the error estimate reflects the increased precision introduced by the pairing process. When a high relationship exists between the matching variable and the criterion variable, this design has the great advantage afforded by the more precise error estimate. However, if some matching variable is chosen which in the end is found to have little or no relationship to the criterion variable, a great deal of time and effort will have been wasted, since no gain in precision will have occurred.

Three or More Paired Groups. This design is simply an extension of that discussed under two paired groups, in which several variations of a variable may be observed in the same investigation. Obviously, as the number of variations is increased, the difficulties of exact matching are also increased. The results from this design are analyzed by the analysis of variance methods.

An alternate technique for increasing precision, which avoids matching in any form but which yields the same degree of precision, is the method of analysis of covariance. By this method the chance inequalities in initial status of group members are adjusted statistically.

One of the disadvantages of designs requiring the matching of subjects is that groups cannot be organized and scheduled until after the initial test has been administered and the subjects matched. When the results are analyzed by the covariance methods, this difficulty is not encountered, since the initial test may be administered after the groups have been organized.

Factorial Designs. Factorial designs make it possible to study two or more variables in the same experiment, each variable having two or more variations. With this kind of design, all possible combinations of the variables and their variations are analyzed.

The 2 $\times$ 2 Factorial Design. This is the simplest of the factorial designs, in which two variables are studied, each with two variations. Four treatment groups are involved, each group being subjected to a different combination of the variations of the variables.

The treatments are organized as follows:

Treatment 1: variation 1 of variable 1 and
variation 1 of variable 2

Treatment 2: variation 1 of variable 1 and
variation 2 of variable 2

Treatment 3: variation 2 of variable 1 and
variation 1 of variable 2

Treatment 4: variation 2 of variable 1 and
variation 2 of variable 2

The subjects available for the experiment are assigned at random to the groups, and the groups are assigned at random to the treatments. At the end of the treatment period, a criterion measure is secured, and the results are analyzed by the methods of analysis of variance.

The great advantage of this design is that there can be derived from a single experiment the same information, with the same degree of precision, that would require two separate experiments under other designs. In addition, information is provided concerning the interaction between the two variables which would be entirely lacking under other designs.

A very good illustration of this design is afforded by the bowling study reported by Summers (31). The two variables of the study are style of delivery and point of aim. The variations of style of delivery are the hook ball and the straight ball; the variations of point of aim are spot and pin. The purposes of the study are to observe the effect of variations in style of delivery and variations in point of aim on the achievement of beginning women bowlers and to determine whether interaction effects occur between delivery and point of aim.

The 72 subjects available for the investigation were assigned at random to four groups, 18 subjects in a group, and the groups were assigned at random to the treatments. The four treatments were (a) hook ball delivery with spot point of aim, (b) hook ball delivery with pin point of aim, (c) straight ball delivery with spot point of aim, and (d) straight ball delivery with pin point of aim. At the end of the instruction period of 12 weeks, the criterion measure was secured. This measure is the cumulative 24-game bowling average with scores adjusted for differences in initial status.

From this single study it was possible to show that the observed difference due to variations in style of delivery could be attributed to chance; that the observed difference in achievement due to variations in point of aim was statistically significant, with the spot point of aim producing the superior result; and that there was no interaction between delivery and point of aim. The meaning of the interaction result is that differences due to variations in point of aim will be the same regardless of which style of delivery is used, and differences due to variations in delivery will be the same regardless of which point of aim is used. If a significant interaction had been found, it would have centered attention on the combinations of delivery and point of aim which would lead to the best achievement.

The 2 × 2 × 2 Factorial Design. This design is an extension of the 2 × 2 design, in which one more variable with two variations is added to the experiment. This would result in eight treatment groups. For example, in the bowling problem a third variable—approach—could be added. The variations of approach could be five steps and four steps. In this case the treatments for the groups would be: (a) hook ball delivery, spot aim, five-step approach; (b) hook ball delivery, spot aim, four-step approach; (c) hook ball delivery, pin aim, five-step approach; (d) hook ball delivery, pin aim, four-step approach; (e) straight ball delivery, spot aim, five-step approach; (f) straight ball delivery, spot aim, four-step approach; (g) straight ball delivery, pin aim, five-step approach; (h) straight ball delivery, pin aim, four-step approach.

This design would make it possible to analyze the effects of variations in style of delivery, the effects of variations in point of aim, and the effects of variations in approach. In addition, it would make possible an analysis of the first-order interaction between delivery and point of aim, delivery and approach, and point of aim and approach; as well as the second-order interaction of delivery, point of aim, and approach.

Complex Factorial Designs. In these designs any number of variables, each with any number of variations, may be studied. As in the simpler factorial designs, the effects of the variations of the main variables are observed as well as the simple and higher order interactions. The main limitations of these more complex designs are in the mechanics of organizing and randomizing subjects across a large number of groups and then of handling these groups. For example, in a 4 $\times$ 3 $\times$ 2 design there would be three main variables, the first variable having four variations, the second three variations, and the third two variations. This design would require the handling of 24 groups, which becomes a particularly limiting factor where the subjects are human beings.

Randomized Blocks Design. One of the limitations common to completely randomized designs is that they frequently fail to provide a desirable degree of accuracy or precision. Alternate designs have been discussed which introduce a greater degree of precision, and the randomized blocks represent still another approach to the solution of this problem.

In this design the subjects of the experiment are arranged into relatively homogeneous groups or blocks and then treatments are assigned at random to the subjects within the blocks. In other words, certain factors which are related to the criterion variable and which are measurable are controlled through the process of developing homogeneous blocks of subjects; other factors not so controlled are randomized.

This design may be illustrated through a simple experiment in which the criterion variable is health knowledge and the experimental treatments are four different films developed to convey certain health knowledges and understandings. The purpose of the investigation is to examine the relative effectiveness of the four films in improving health knowledge. The educational level of the films is junior high school, so an experiment is planned for children at that level.

Twenty-four children are to be used in the investigation. Under a completely randomized design, the children would be assigned at random to four groups and the four groups would be assigned at random to the four treatments. In the randomized blocks design, however, the investigator would attempt to identify certain attributes of the children which would be relevant to their response to the experimental treatments. For example, sex might be a factor to be considered. Assume that of the 24 children, 12 are males and 12 are females. The subjects would be divided into two groups or blocks of 12 each, according to their sex. Within each block the four treatments would be assigned at random to the subjects, with three subjects in each block receiving each treatment. The comparisons between treatments would then be for children of the same sex. In a design which is completely randomized, the number of children of each sex being subjected to the various treatments may be unequal just by chance and the resulting differences in the criterion measure may very well be at least partially the consequence of sex differences.

A second factor which doubtless would be a consideration in this investigation is intelligence. The children in each of the sex blocks could be divided, for example, into three intelligence levels, resulting finally in six blocks with four children in each block. Three of these blocks would be males, one of the three would contain the four males with the highest intelligence, one would contain the four males of medium intelligence, and the third the four males of lowest intelligence. The same arrangement, by intelligence, would be established for the 12 females. Finally, within each block, the four treatments would be assigned at random to the four subjects. The comparisons in this arrangement would be for children of the same sex and relatively homogeneous intelligence.

These same techniques could be continued to include as many equalizing factors as desired, although as the number of factors is increased so must the number of subjects be increased. This design permits any number of treatments and any number of subjects under each treatment. The results are analyzed by the methods of analysis of variance.

BIBLIOGRAPHY

1. Ackoff, Russell L. *The Design of Social Research.* Chicago: University of Chicago Press, 1953.
2. Andrews, T. G., editor. *Methods of Psychology.* New York: John Wiley & Sons, 1948.
3. Barr, A. S.; Davis, R. A.; and Johnson, P. O. *Educational Research and Appraisal.* Chicago: J. B. Lippincott Co., 1953.
4. Bartley, S. Howard. *Beginning Experimental Psychology.* New York: McGraw-Hill Book Co., 1950.
5. Bogardus, Emory S. *Introduction to Social Research.* Los Angeles: Kellaway-Ide-Jones Co., 1936.
6. Chapin, F. Stuart. *Experimental Designs in Sociological Research.* New York: Harper & Brothers, 1947.
7. Cochran, William G., and Cox, Gertrude. *Experimental Designs.* Second edition. New York: John Wiley & Sons, 1950.
8. Edwards, Allen L. *Experimental Design in Psychological Research.* New York: Rinehart & Co., 1950.
9. Elmer, M. C. *Social Research.* New York: Prentice-Hall, 1939.

10. FESTINGER, LEON, and KATZ, DANIEL, editors. *Research Methods in the Behavioral Sciences.* New York: Dryden Press, 1953.
11. FISHER, SIR RONALD A. *The Design of Experiments.* Sixth edition. Edinburgh: Oliver and Boyd, Ltd.; New York: Hafner Publishing Co., 1951.
12. FISHER, SIR RONALD A. *Statistical Methods and Scientific Inference.* Edinburgh: Oliver and Boyd, Ltd.; New York: Hafner Publishing Co., 1956.
13. FISHER, SIR RONALD A., and YATES, F. *Statistical Tables for Biological, Agricultural and Medical Research.* Fourth edition. Edinburgh: Oliver and Boyd, Ltd.; New York: Hafner Publishing Co., 1953.
14. GARRETT, HENRY E. *Great Experiments in Psychology.* Third edition. New York: Appleton-Century-Crofts, 1951.
15. GEE, WILSON. *Social Science Research Methods.* New York: Appleton-Century-Crofts, 1950.
16. GOOD, C. V.; BARR, A. S.; and SCATES, D. E. *The Methodology of Educational Research.* New York: Appleton-Century-Crofts, 1936.
17. GOOD, C. V., and SCATES, D. E. *Methods of Research.* New York: Appleton-Century-Crofts, 1954. (See also GOOD, C. V. *Introduction to Educational Research.* 1959.)
18. GOODE, W. J., and HATT, P. K. *Methods in Social Research.* New York: McGraw-Hill Book Co., 1952.
19. JOHNSON, PALMER O. *Statistical Methods in Research.* New York: Prentice-Hall, 1949.
20. KEMPTHORNE, OSCAR. *The Design and Analysis of Experiments.* New York: John Wiley & Sons, 1952.
21. LEV, JOSEPH. "Research Methods and Designs." *Review of Educational Research* 18: 410-23; December 1948.
22. LINDQUIST, E. F. *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton-Mifflin Co., 1953.
23. LUNDBERG, GEORGE A. *Social Research.* New York: Longmans, Green & Co., 1942.
24. MADGE, JOHN. *The Tools of Social Science.* New York: Longmans, Green & Co., 1953.
25. MANN, H. B. *Analysis and Design of Experiments.* New York: Dover Publications, 1949.
26. MORENO, J. L. *Sociometry, Experimental Method and the Science of Society.* New York: Beacon House, 1951.
27. NORTON, DEE W., and LINDQUIST, E. F. "Applications of Experimental Design and Analysis." *Review of Education Research* 21: 350-67; December 1951.
28. OSGOOD, CHARLES E. *Method and Theory in Experimental Psychology.* New York: Oxford University Press, 1953.
29. STEVENS, S. S., editor. *Handbook of Experimental Psychology.* New York: John Wiley & Sons, 1951.
30. SUMMERS, DEAN. *The Effect of Variations of Type Delivery and Type Aim on Bowling Achievement of College Women.* Bloomington: Indiana University. Microcard PE 281. Also, *Research Quarterly* 28: 77-84; March 1957.
31. TINKER, MILES A. *Introduction to Methods in Experimental Psychology.* Second edition. New York: Appleton-Century-Crofts, 1947.
32. TOWNSEND, JOHN C. *Introduction to Experimental Method: For Psychology and the Social Sciences.* New York: McGraw-Hill Book Co., 1953.
33. UNDERWOOD, BENTON J. *Experimental Psychology: An Introduction.* New York: Appleton-Century-Crofts, 1949.
34. VALENTINE, WILLARD L., and WICKENS, DELOS D. *Experimental Foundations of General Psychology.* Third edition. New York: Rinehart & Co., 1949.
35. VISSCHER, MAURICE B., editor. *Methods in Medical Research.* Chicago: Year Book Publishers, 1951. Vol. 4.
36. YOUNG, PAULINE V. *Scientific Social Surveys and Research.* Second edition. New York: Prentice-Hall, 1949.

CHAPTER 11

Laboratory Research

MARGARET FOX
C. ETTA WALTERS
JOHN COOPER
OLIVE YOUNG
M. GLADYS SCOTT
MARJORIE U. WILSON
JOSEF BROZEK, Consultant

THE TERM LABORATORY IS USED IN THIS CHAPTER TO ENCOMPASS all those sites where research of an applied type takes place. It may be a room especially equipped for particular kinds of experimentation as cited in Chapter 6. It may be the gymnasium, dance studio, swimming pool, or athletic field.

The discussions which follow deal with certain areas of applied research, each of which has more or less implication for the fields of health, physical education, and recreation. These areas are discussed here to show, at least in part, what has been done, the problems, and the limitations and needs. Although they have certain similarities and interrelationships to which the reader should be alert, they are presented here as separate discussions for reasons of expediency.

Anthropometry

MARGARET FOX

Although it may be said of all studies that the value of the work depends on the meticulousness with which subjects are measured, it is especially true of anthropometric studies. Care needs to be taken not only in defining body landmarks precisely but also in selecting an instrument for measuring which will not be subject to variation.

It is equally true that the population limits need to be precisely set, since many anthropometric studies concern themselves with describing differences in body growth and development. These, in turn, are dependent upon many factors, such as geographical location, race, age, and body type. Even these factors can be broken down into finer units, such as altitude and temperature variations; country of predecessors; chronological, skeletal, or physiological age; and endomorphic, mesomorphic, or ectomorphic body types.

But even with satisfactorily defined landmarks, adequate instrumentation, and population limitations, the study will be worthless if adequate attention is not paid to the method of taking and recording measurements. Variance caused by avoidable errors can accumulate to a surprising amount if all these factors are not taken into consideration.

PROBLEMS OF ANTHROPOMETRIC MEASUREMENT

Landmarks. While anthropologists tend to be precise in describing body landmarks, much of the professional literature in physical education continues to describe anatomical boundaries in rather vague terms. Fox and Young (12) reveal this lack of precision in a survey of early studies concerning the points at which the line of gravity intersects the knee and ankle joints. These authors attempt to define the exact spot of intersection with greater accuracy.

Another landmark often inadequately described is the acromion process. Investigators may attempt to clarify it by specifying the tip of the acromion. This scarcely adds any light, for the most lateral part of the acromion measures more than an inch across. Does the investigator have in mind the most lateral projection of the acromion or the anterior or the posterior angle of the acromion? In trying to locate a gravital line in relation to such a loosely described landmark there may be a variation of as much as an inch, solely the result of vague description of the landmark.

In measuring height, it is necessary not only to be certain that the hair does not interfere with the measurement but also to observe whether the head is held on the "Frankfort Plane," or variations of fractions of inches may occur.

Another problem in selecting a landmark is the ease with which such a point can be located. Subcutaneous fat or muscle mass may make it difficult to find the landmark. If there is a choice available, the landmark nearest the surface and with the least amount of soft tissue interference should be chosen.

Position and Condition of the Subject. Position and condition of the subject may lead to avoidable error in taking anthropometric measurements. For example, considerable difference can be noted in the sitting height of the individual according to the position of the legs. If the knees are together, each bent at a right angle, and the feet are flat on the floor, the gluteal muscle mass and, to a certain extent, the tensing of the muscles add to the height of the individual. If the ankles are crossed and the knees separated, the muscles relax and there is a closer approximation of the true skeletal stem length from the tuberosities of the ischia to the vertex of the head. Muscle tension will influence girth measurements as well as sitting heights. Standardization of muscle tension as well as of position will be necessary to minimize this source of error.

Height and weight vary during the day. Individuals not only tend to be taller early in the day but also to weigh less. If considerable precision is desired in taking records of this type, the time factor must be considered. Proximity of meals and the timing of bowel and bladder evacuation may have marked effect on weight of young subjects but will be of less consequence to adult measurements.

Variations in posture may have their effects. One difficulty in trying to measure shoulder width is the position of the shoulder girdle, while a fatigue slump may affect the total height. These errors can be minimized to a certain extent by the directions given by the examiner.

Needless to say, shoes and other clothing may cause marked differences in height and weight because of the variations in wearing apparel from individual to individual. As a standard procedure, subjects should remove their shoes before measurements are taken.

In longitudinal studies where height is a factor, it may be necessary to take some measurements in the reclining position and some in the standing position because of the age of the subjects. Researchers do not agree upon the age at which the shift from one position to another should be standardized and at which the position mentioned in reporting results should be listed as standing or reclining height.

Measurements of the thorax vary considerably if taken on inspiration or expiration, or if taken when the individual has the chest elevated, as when looking forward, or depressed, as when looking down.

Accuracy. There are a number of factors which may influence the accuracy of the records even when sufficient care is taken in the choice of landmarks, instruments, and position of the subject.

Of prime importance is the skill of the anthropometer. He may need considerable practice in taking the selected measurements in order to duplicate his own results with any degree of reliability. This also may be dependent on his physical condition, because fatigue, inability to see well, lack of ability to concentrate, and/or boredom may nullify all the work of the designer of the experiment.

The pressure with which instruments are applied must be specified before measurement begins and must be uniform from subject to subject when data are taken. The instruments themselves must not vary with weather conditions or use; tapes may shrink and springs may lose their tension. Instruments should be calibrated from time to time to avoid these errors.

Planning the order of observations is necessary to avoid having to change instruments too often or to vary the position of the observer too frequently. Figures need to be carefully recorded so there is no subsequent question of whether a "1" or "7" was intended. The aid of a competent clerk who can pay attention to verbal directions is desirable. The use of such a person acts as a check on the observer as to the feasibility of such an observation. A recorder also makes it unnecessary for the anthropometer to jot down each figure as it is observed for fear that he will make a mistake in attempting to keep in mind several measurements taken in succession.

Units of Measure. The English-speaking countries are plagued with having to adjust to two sets of measurements. Much of our scientific measurement is in the metric system. However, many reports for lay usage are still in terms of inches, feet, and pounds (26, 27). While it is customary to report fractions of an inch in quarters, eighths, and so forth, this division is not practical from the standpoint of computations. It is easier to add tenths of inches, when necessary, to determine means and standard deviations. This means that special instruments calibrated in that fashion must be used. The metric system obviates this difficulty.

The size of the measuring unit is of major importance. In measuring the total height of an individual, it may be sufficient to measure to the nearest half inch or the nearest centimeter. But in measurements such as ankle and wrist girth, the nearest tenth of an inch or half of a centimeter is indicated.

For metric measures, Tildesey (51) recommends that, when the standard deviation is less than 3.3 mm, the unit of measurement should be smaller than a millimeter. If the standard deviation is less than 33 mm, she recommends that the unit of measurement should be less than a centimeter. Her recommendations are based on statistical evidence that, for direct measurement, the unit should not be more than two-thirds of the standard deviation of the character being measured. It is customary in most direct measures in physical education to use the centimeter, except when a derived measure is to be used, such as computing the leg length by subtracting the sitting height from the standing height. In this derived measure, the millimeter should be the unit of measure.

SELECTION OF SUBJECTS

There are a number of factors which should be taken into account in selecting subjects for any anthropometric study. Setting the population limits so that a relatively homogeneous group can be obtained may yield more consistent results. There are a number of studies which make this observation apparent.

In a review of 25 studies, Kaplan (20) has concluded that climate, diet, and altitude significantly influence growth patterns. Newman and Munro (34) also note that there is a correlation between colder climates and greater weight and surface area of army inductees. Their suggested explanation is that colder temperatures stimulate activity and appetite. Roberts (43) also noted this correlation. In a study involving Oregon children, the Merediths (30) noted that they were taller and heavier and had greater girths than children from a number of areas of the country with a similar ethnic background. These studies suggest that there are real population differences attributable to climate.

Race must also be considered. Trotter and Gleser (52) have found that Negroes have significantly longer bones in the extremities than subjects from the white race. Other observations from their study indicate that there is a decrease in stature due to aging.

In the selection of subjects, therefore, it would be well to narrow the population and increase the number of subjects for consistency of results. With such marked geographic and racial differences, it should be apparent that in physical performance events where height and weight might have some influence it is futile to attempt to set national norms in this country.

PROBLEMS STILL TO BE SOLVED

A number of problems still remain to be answered by research in anthropometry. For example, no satisfactory method has yet been worked out for classifying women by body types. Apparently there is a sex difference in the important components making up the classifications. Until a method of classifying women by body type has been worked out, there can be little research on the relationship of body type to physical performance of women.

Somatotyping for men would bear simplification. In addition, a longitudinal study of somatotypes, particularly as they affect the middle and older age groups, is indicated. The relationship of the somatotype to various constitutional diseases would be of value in preventive medicine.

Further study needs to be made of racial differences in skeletal and muscle structure, in order to give a clue to the superior performance by the Negro race in some types of activities. A study of relative buoyancy of Negroes and whites would be valuable, as well as a study of postural differences between the races.

The relationship of climate to growth has been studied to a limited extent, but this would bear verification and extension.

Although some work has been done on the relation of skeletal maturity to physical performance, further investigation of skeletal maturity in various parts of the body should be carried out.

It is believed that muscles develop at various rates, but little is known of the relationship of skeletal and muscular maturity or at what times one can expect certain muscle groups to have developed. Another unknown is the value of specified activities to promote earlier development of muscles.

The relationship of muscle size to strength has been worked out in the laboratory but bears investigation in the living subject. Can certain types of exercises increase strength without causing marked hypertrophy of muscles? What is the mathematical relationship between growth in strength and hypertrophy of muscles?

A number of problems have been suggested that could be solved by anthropometric research.

STUDIES RELATED TO BODY BUILD

In recent years there have been a number of methods of ascertaining body build. These are of interest to the physical educator because there are indications that, at least for males, performance is related to build. Carpenter's work (5, 6) with women suggests that the body-build factor is not so important with that sex, however.

One of the early classifications of physical performance by age, height, weight is McCloy's (25). He found that age alone was sufficient for girls. This type of classification is most satisfactory for groups below college age when growth is still taking place. Miller (31) concluded that height and weight were unsatisfactory elements on which to base classification of college men.

More recent methods include that of Sheldon's classification by somatotype (45). Photographs are taken of the subjects under standardized conditions. The predominant component of physique is judged by comparing the photograph with various standards pictured in his *Atlas*. Sheldon has worked out three major components which he has labeled *endomorphy* or the tendency toward roundness of body form; *mesomorphy* or the predominance of muscle, bone, and connective tissue; and *ecto-*

morphy or the tendency toward fragility and linearity. Presumably these components form a continuum, and a physique may contain elements of all three components. Sheldon describes 88 types, with a total of 505 types when half scores are used.

Both proponents and opponents of this method will be found. Lorr and Fields (23), in a factorial analysis of body types, found three distinguishable groups closely resembling Sheldon's. Sills (46) describes a fourth component, omomorphy, in the classifications. The latter component is characterized by a V-shaped torso with the chest and shoulders more highly developed than the lower extremity. He also found that motor ability tests had positive relationships to mesomorphy and omomorphy, while strength did not have a significant relation to any of the components. Sills and Everett (47) suggest that consideration should be given to body types such as Sheldon describes in formulating standards for achievement on strength or motor tests because mesomorphs were superior to the other two types in such tests. Sills and Mitchem (48) grouped 13 classifications of somatotypes into four groups and then computed T-scores for predicted performance on physical fitness tests for the groups.

Few studies have been done where women have been classified by somatotypes. Though Sheldon and his co-workers have been studying the application of this technique to women, they have not published their results to date. Perbix (38) studied the relation of somatotype ratings to motor fitness of college women. She concluded that her groups as a whole had endomorphy as their dominant component but that physical education majors tended to be more dominant in mesomorphic traits. She also found a significant relation between mesomorphy and knee push-ups and medicine ball put.

In a longitudinal study, Dupertius and Michael (9) compared growth in height and weight between ectomorphic and mesomorphic boys and found that over a period from 4 to 17 years of age the ectomorphs were consistently taller while the mesomorphs were heavier. The ectomorphs grew over a longer period, while the mesomorphs grew faster and matured earlier. They believe that somatotypes, as far as ectomorphs and mesomorphs are concerned, remain fairly constant throughout childhood and young adulthood.

It seems apparent that there is a heredity core of parent-child similarities in physical characteristics, according to Bayley (1). This likeness becomes more pronounced as children mature. Peckos' study (37) substantiates the hereditary feature of body-build inheritance in her study on caloric intake of children. She found that endomorphic children had the lowest caloric intake of the three classes, while ectomorphic children had the highest. She concluded that body build was not a function of caloric intake alone.

The practice of taking measurements of body build from photographs has been questioned by some investigators, but Geoghegan (14), comparing actual measurements with those computed from photographs taken under standard conditions, concludes that measurements taken from photographs prove satisfactory.

There appears to be enough evidence that when physical performance standards are set for boys some consideration should be given to body build. The situation for girls is not so clear-cut, and investigation of this problem seems desirable.

TRENDS IN PHYSICAL GROWTH

In the past few years, there has been increased interest in growth and development of the child, particularly as they pertain to the physical fitness of the individual. Much of the early work had to do with collecting data on range in size of various parts of the body at various age levels. Other investigators have collected data for use of architects and school designers (26, 27). Tuddenham and Snyder (53), using a longitudinal design, collected data not only on growth in body size but also in strength over the period of birth to 18 years of age.

There continues to be interest in assessing maturity by means of observations of osseous development (16, 17, 36). Nicolson and Hanley (35) describe measures of physical maturity, giving age norms for the appearance of these signs of maturity.

Of interest to the physical educator are observations on the relation of growth and maturity to the development of strength. Weinback (55) states that increase in strength in children is not proportional to increase in size. This observation is significant in such activities as tumbling, where the inexperienced teacher is apt to choose the largest or tallest children as bases in couple and group stunts under the mistaken impression that they are the strongest.

In relation to body build and growth, Steinhaus (49) observes that mesomorphs build more muscle than linear types when given the same amount of exercise. Jones (19) notes that while static strength is associated with body size in early adolescence it is more closely related to body build. For the same body weight, mesomorphs have more strength. In relation to maturity, Jones also notes that eight-year-old mesomorphic boys are closer to sexual maturity and stronger than eight-year-old ectomorphs. He suggests, as a result of his observations, that teachers should give consideration to the constitutional make-up of children in establishing levels of expectation.

Rarick (39) and Jensen (18) have synthesized some of the pertinent findings of studies done on growth. Reynolds (41) noted that girls who reach sexual maturity at an early age show a spurt in the growth of the muscle mass of the leg at 9½ years, while late maturing girls show this spurt 2½ years later. The boys' spurt is similar in pattern to the girls but comes later. He also observed that there was a sharp increase in leg extensor strength at 11 years of age. This observation seems pertinent to the selection of activities for elementary school boys.

Wetzel (56, 57), who was interested in growth failure as a method of locating children with incipient health problems, devised a grid for plotting growth of children. Several studies have been done observing the functioning of the Wetzel grid and using it as a method of classifying children for motor activities. Watson and Lowrey (54) discuss the grid in their study on child growth. Krogman (21) gives directions for using it.

Although the Wetzel grid was planned primarily for assessing growth of children, Miller (32) used it as a classification device in studying the performance of college men on motor activities. He concluded that the grid had dubious value as a performance classifier for college men for motor performance. Garn (13) applied the grid technique in a longitudinal study of girls but found such a number of deviations from the starting channel within two years that he believes that constancy of channel position is not usual for girls during growth and maturity.

Hall (15) plotted data from observations of growth of 4-H club members in Illinois on several types of growth indexes.

He also gives growth curves of several body measurements for both boys and girls in the study.

There has been some interest in subcutaneous fat measurements in relation to growth, body temperature, and basal metabolism rates for boys and girls by Reynolds (42) and Eichorn and McKee (10).

It has been suggested that there is an interrelationship of the various growth measures such as height, weight, strength, osseous development, vital capacity, dental development, and reading development. The composite has been termed the *organismic age*. Blommers and others (3) found no systematic tendency for these various "ages" to be related except by chance.

One of the major difficulties in studying growth and development has been that of carrying out longitudinal studies. Bell (2) describes a plan for an accelerated longitudinal approach which he terms convergence. This procedure should prove useful to at least part of the studies and is well worth investigating as a possibility when planning this type of study.

One of the simpler and more feasible methods of measuring growth in children is that proposed by Meredith (28, 29). He has prepared a booklet, *Physical Growth Record,* in which the growth curve may be plotted for children from 4 to 18. There are separate forms available for each sex. In the interior of the booklet, five growth zones are separately plotted for height and weight. The weight or height is plotted in a zone which corresponds to the intersection of two lines. Age is plotted by half-year intervals on the vertical lines, while height or weight is plotted on horizontal lines. The two lines will intersect in a growth zone. By comparing height and weight zones for consistency, marked discrepancies can be determined and such cases referred to a physician for assessment of physical status. This method has the advantage of simplicity and of economy from the standpoint of time.

Curriculum planners would do well to consider some of the findings on growth and development in planning curricular changes. As in body build, while many of these methods function reasonably well for males, considerable study needs to be done on how these methods apply to females.

APPLICATION OF ANTHROPOMETRIC DATA

Physical and health educators are apt to be interested in anthropometry as it pertains to the solution of some problem in the field. For example, what is the relation of length of body segments to flexibility? Does variation in height seriously influence performance in tests in the sports area? Should body build be used to classify students for physical fitness tests? The few examples to be given here may stimulate investigators to make greater use of these valuable tools.

In times of national emergency, fitness becomes increasingly important. There is always the problem of setting standards of performance, but should the same standard be expected of all? Loveless' study (24) of performance of varying age groups on war-time Navy Physical Fitness Tests indicated that age makes little difference in those from 17 to 30 years old but that those over 30 made consistently lower scores. He also found that height had little influence on the scores but that weight over 190 pounds did cause a decrease in scores.

In the area of sports, Lamp (22) found that at the junior high level sex differences in volleyball skill were insignificant. However, age and weight were significantly related to volleyball skill for girls, while height was for boys. Maturity also was definitely related to performance by boys.

An investigation of relationships of age, height, and weight to track and field performance by Cearley (7) produced the conclusion that performance ability of both boys and girls 9 to 17 years old in track and field events bears a nonlinear relation to age, height, and weight. Age made its greatest contribution to performance in that area at 15½ years for boys and 13½ years for girls. Height made its greatest contribution at 71 inches for boys and 51 inches for girls, while weight made its greatest contribution for both boys and girls at 55 pounds.

In a study on college women, Mohr and Haverstick (33) found height to have some relationship to volleying ability as demonstrated in the volleyball wall volley when the subject stood three feet from the wall. DiGiovanna (8) concluded from his investigation that body structure and muscular strength are associated with athletic success. Everett and Sills (11) noted that weight, anthropometric measurements of the hand, height, and mesomorphy all correlated with grip strength.

Seils (44) did an extensive study on the relationship of measures of physical growth and gross motor performance of primary grade children. Mean performance in the skill events tended to increase with the increase in age. Rarick (40) was also interested in primary grade children and the relation of growth to strength. His findings are that at the seven-year-old level there are qualitative sex differences in muscle tissue as measured by roentgenograms. In addition, he concluded that boys have greater muscular power per unit of muscle mass than girls.

In studying the relationship of body size and shape to physical performance, Bookwalter (4) found that maximum size and shape do not produce maximum physical fitness and that the thin and average perform equally well physically. He concluded there was a systematic relationship between developmental levels and fitness scores.

Tanner (50) studied the effect of four months of weight training on the physique of ten mesomorphs. After the subjects had a four-month rest, he found that all measurements, with the exception of the upper arm girth, had reverted to normal pretraining size. As a result, he believes the muscle growth potential of the arms is greater than that of the legs.

These studies point the way toward additional research linking body size and maturity with performance in gross motor activities.

BIBLIOGRAPHY ON ANTHROPOMETRY

1. Bayley, Nancy. "Some Increasing Parent-Child Similarities During the Growth of Children." *Journal of Educational Psychology* 45:1-21; January 1954.
2. Bell, Richard Q. "Convergence: An Accelerated Longitudinal Approach." *Child Development* 24:145-52; June 1953.
3. Blommers, Paul; Knief, Lotus M.; and Stroud, J. B. "The Organismic Age Concept." *Journal of Educational Psychology* 46:142-50; March 1950.
4. Bookwalter, Karl W. "The Relationship of Body Size and Shape to Physical Performance." *Research Quarterly* 23:271-79; October 1952.
5. Carpenter, Aileen. "An Anthropometric Study of Masculinity and Femininity of Body Build." *Research Quarterly* 12:712-19; December 1941.
6. Carpenter, Aileen. "Strength, Power, and 'Femininity' as Factors Influencing the Athletic Performance of College Women." *Research Quarterly* 9:120-27; May 1938.
7. Cearley, Jess E. "Linearity of Contributions of Ages, Heights, and Weights to Prediction of Track and Field Performances." *Research Quarterly* 28:218-22; October 1957.
8. DiGiovanna, Vincent. "The Relationship of Selected Structural and Functional Measures to Success in College Athletics." *Research Quarterly* 14:199-216; May 1943.

9. DUPERTUIS, C. WESLEY, and MICHAEL, NANCY B. "Comparison of Growth in Height and Weight between Ectomorphic and Mesomorphic Boys." *Child Development* 24:203-14; September 1953.
10. EICHORN, DOROTHY H., and MCKEE, JOHN P. "Oral Temperature and Subcutaneous Fat During Adolescence." *Child Development* 24:235-47; September 1953.
11. EVERETT, PETER W., and SILLS, FRANK D. "The Relationship of Grip Strength to Stature, Somatotype Components and Anthropometric Measures of the Hand." *Research Quarterly* 23:161-66; May 1952.
12. FOX, MARGARET G., and YOUNG, OLIVE G. "Placement of the Gravital Line in Antero-Posterior Standing Posture." *Research Quarterly* 25:277-85; October 1954.
13. GARN, STANLEY M. "Individual and Group Variations from 'Channelwise' Grid Progression in Girls." *Child Development* 23:193-206; September 1952.
14. GEOGHEGAN, BASIL. "The Determination of Body Measurements, Surface Area, and Body Volume by Photography." *American Journal of Physical Anthropology* 11:97-119; March 1953.
15. HALL, D. M. "Determining Healthy Growth for 4-H Club Members." *Research Quarterly* 24:284-94; October 1953.
16. HARDING, VERNETTE V. "A Method of Evaluating Osseous Development from Birth to 14 Years." *Child Development* 23:247-71; December 1952.
17. HARDING, VERNETTE V. "Time Schedule for the Appearance and Fusion of a Secondary Accessory Center of Ossification of the Calcaneus." *Child Development* 23:181-84; September 1952.
18. JENSEN, KAI. "Physical Growth." *Review of Educational Research* 25:369-415 December 1955.
19. JONES, HAROLD E. *Motor Performance and Growth.* Berkeley: University of California Press, 1949.
20. KAPLAN, BERNICE A. "Environment and Human Plasticity." *American Anthropologist* 56:780-800; February 1953.
21. KROGMAN, WILTON M. *A Handbook of the Measurement and Interpretation of Height and Weight in the Growing Child.* Monographs of the Society for Research in Child Development, 18:3. Evanston: Child Development Publications, 1950.
22. LAMP, NANCY A. "Volleyball Skills of Junior High School Students as a Function of Physical Size and Maturity." *Research Quarterly* 25:189-200; May 1954.
23. LORR, MAURICE, and FIELDS, VICTOR. "A Factorial Study of Body Types." *Journal of Clinical Psychology* 10:182-85; 1954.
24. LOVELESS, JAMES C. "Relationships of the War-Time Navy Physical Fitness Test to Age, Height, and Weight." *Research Quarterly* 23:347-55; October 1952.
25. MCCLOY, CHARLES H. *Tests and Measurements in Health and Physical Education.* Third Edition. New York: Appleton-Century-Crofts, 1954.
26. MARTIN, W. EDGAR. *Basic Body Measurements of School Age Children.* U.S. Department of Health, Education, and Welfare, Office of Education, Washington, D.C.: Superintendent of Documents, Government Printing Office, June 1953.
27. MARTIN, W. EDGAR. *The Functional Body Measurements of School Age Children.* Chicago: National School Service Institute. 1954.
28. MEREDITH, HOWARD V. *Physical Growth Records for Boys/Girls,* Washington, D.C.; and Chicago: Joint Committee on Health Problems in Education of the National Education Association and the American Medical Association, 1947. 4 p.
29. MEREDITH, HOWARD V. "A 'Physical Growth Record' for Use in Elementary and High Schools." *American Journal of Public Health* 39:878-85; July 1949.
30. MEREDITH, HOWARD V., and MATHILDA, E. "The Body Size and Form of Present-Day White Elementary Children Residing in West-Central Oregon." *Child Development* 24:83-102; June 1953.

31. MILLER, KENNETH D. "A Critique on the Use of Height-Weight Factors in the Performance Classification of College Men." *Research Quarterly* 23:402-16; December 1952.
32. MILLER, KENNETH D. "The Wetzel Grid as a Performance Classifier with College Men." *Research Quarterly* 22:63-70; March 1951.
33. MOHR, DOROTHY R., and HAVERSTICK, MARTHA. "Relationship Between Height, Jumping Ability, and Agility to Volleyball Skill." *Research Quarterly* 27:74-78; March 1956.
34. NEWMAN, RUSSELL W., and MUNRO, EDNA H. "The Relation of Climate and Body Size in U.S. Males." *American Journal of Physical Anthropology* 13:1-17; March 1955.
35. NICOLSON, ARLINE B., and HANLEY, CHARLES. "Indices of Physiological Maturity: Derivation and Interrelationships." *Child Development* 24:3-38; March 1953.
36. NOBACK, CHARLES R. "The Appearance of Ossification Centers and the Fusion of Bones." *American Journal of Physical Anthropology* 12:63-69; March 1954.
37. PECKOS, PENELOPE S. "Caloric Intake in Relation to Physique in Children." *Science* 117:631-33; June 1953.
38. PERBIX, JOYCE A. "Relationship Between Somatotype and Motor Fitness in Women." *Research Quarterly* 25:84-90; March 1954.
39. RARICK, LAWRENCE. "Maturity Indicators and the Development of Strength and Skill." *Education* 75:69-73; October 1954.
40. RARICK, LAWRENCE, and THOMPSON, JOANN JONES. "Roentgenographic Measures of Leg Muscle Size and Ankle Extensor Strength of Seven-Year-Old Subjects." *Research Quarterly* 27:321-32; October 1956.
41. REYNOLDS, EARLE L. "Differential Tissue Growth in the Leg During Childhood." *Child Development* 15:181-205; December 1944.
42. REYNOLDS, EARLE. *"The Distribution of Subcutaneous Fat in Childhood and Adolescence."* Child Development Publications of Society for Research in Child Development. Evanston, Ill.: Northwestern University, 1951.
43. ROBERTS, D. F. "Body Weight, Race, and Climate." *American Journal of Physical Anthropology* 11:533-58; December 1953.
44. SEILS, LEROY G. "The Relationships Between Measures of Physical Growth and Gross Motor Performance of Primary-Grade School Children." *Research Quarterly* 22:244-60; May 1951.
45. SHELDON, W. H.; STEVENS, S. S.; and TUCKER, W. B. *The Varieties of Human Physique.* New York: Harper & Brothers, 1940.
46. SILLS, FRANK D. "A Factor Analysis of Somatotypes and of Their Relationships to Achievement in Motor Skills." *Research Quarterly* 21:424-37; December 1950.
47. SILLS, FRANK D., and EVERETT, PETER W. "The Relationship of Extreme Somatotypes to Performance in Motor and Strength Tests." *Research Quarterly* 24:223-28; May 1953.
48. SILLS, FRANK D., and MITCHEM, JOHN. "Prediction of Performance on Physical Fitness Tests by Means of Somatotype Ratings." *Research Quarterly* 28:64-71; March 1957.
49. STEINHAUS, ARTHUR H. "Chronic Effects of Exercise." *Physiological Reviews* 8:103-107; 1933.
50. TANNER, J. M. "The Effect of Weight Training on Physique." *American Journal of Physical Anthropology* 10:427-60; December 1952.
51. TILDESEY, M. L. "Choice of the Unit of Measurement in Anthropometry." *Man* 47:72-77; May 1947.
52. TROTTER, MILDRED, and GLESER, GOLDINE C. "Estimation of Stature from Long Bones of American Whites and Negroes." *American Journal of Physical Anthropology* 10:463-514; December 1952.
53. TUDDENHAM, READ D., and SNYDER, MARGARET M. *Physical Growth of California Boys and Girls from Birth to Eighteen Years.* University of California Publications in Child Development 1:183-364. Berkeley: University of California Press, 1954.

54. WATSON; ERNEST H., and LOWREY, GEORGE H. *Growth and Development of Children.* Chicago: Year Book Publishers, 1952. p. 44-76.
55. WEINBACH, A. P. "The Human Growth Curve: II. Birth to Puberty." *Growth* 5:235-55; 1941.
56. WETZEL, NORMAN C. *Instruction Manual in the Use of the Grid for Evaluating Physical Fitness.* Cleveland, Ohio: Newspaper Enterprises Association Service, 1941.
57. WETZEL, NORMAN C. *Treatment of Growth Failure in Children.* Cleveland, Ohio: Newspaper Enterprises Association Service, 1948.

Body Mechanics

MARGARET FOX

Few areas in physical education have been taught for as long a period of time as posture, yet the field of posture and body mechanics has been relatively little explored by research methods. Improvement in posture and physique was the major objective of the early work in physical education. Yet, after more than half a century, most of the teaching in this area is still based on empirical judgment.

Interest in research in this area has declined in the past few years. During the period 1930-39, 33 articles were published in the *Research Quarterly* under the subject heading of posture. From 1940-49, only four more additions were made to the literature in that field in the *Research Quarterly,* and since 1949 very few such studies have appeared in the *Quarterly.* A survey of medical journals also indicates little interest in this area, except in foreign literature. Although most of the reports that are available are not of a research nature, this does not mean that all the problems are solved. On the contrary, work has hardly begun.

PROBLEMS IN POSTURE MEASUREMENT

Some of the early work concentrated on the center of gravity and its importance to posture. Reynolds and Lovett (23), Cureton and Wickens (3), Elftman (4), and Hellebrandt and her associates (8, 9, 11) worked out many of the techniques for studying this problem. Fox and Young (5) applied some of their techniques in determining the specific placement of the gravital line in the ankle joint and the knee. Additional work needs to be done in this area in determining the placement of the gravital line in relation to the joints above the knee in optimum posture. Does the gravital line in antero-posterior standing posture run anterior to the head of the femur, through it, or posterior to it? In normal subjects, especially women, it is difficult to determine landmarks at the hip joint, so it is probable that determination of the gravital line would necessitate use of simultaneous roentgenograph and center of gravity readings similar to the method used by Fox and Young. The body segments above the hip also need investigation relative to the placement of the gravital line. What specific body landmarks should be in vertical alignment?

Posture Tests. Having determined where the gravital line should lie in the individual with the optimum stance posture, the next logical step is the devising of some workable test for measuring deviations from that position. Although considerable effort has been expended on the problem of measuring deviations, no completely satisfactory method has yet been found. Several objective tests are available. However, there is usually some objection to each of them.

Kellogg's test (16), which was one of the earliest, has been used as a base for later studies. His method was based on a series of angular measurements, but because it is an early study the standards are inadequate and incomplete. Using a method of angular measurements similar to those of Kellogg, Massey (19) set up a test which he found reliable and valid. However, it has been difficult to duplicate his results on other groups.

MacEwen and Howe (18) approached the problem from another angle, that of measuring the depth of spinal curves. The reliability of their method varied, depending on whether duplicate or successive pictures were graded or on whether the entire process of preparing the subject and taking the picture was repeated. Although the method was carefully validated on a single picture basis, Hellebrandt (7, 10) found that body sway seriously affected reliability when pictures in a series were compared. Low reliability would affect the validity of the test. The MacEwen-Howe test also had the disadvantage of requiring time to prepare materials for testing and for grading test results.

Cureton, in collaboration with others (3), tried to overcome these last objections by using the Conformateur and small electric lights taped to the skin to indicate landmarks. Nine measurements of segmental position, as well as center of gravity determinations and foot measurements, were made by this method.

Despite these carefully planned tests, there is still no one test which satisfies test criteria for antero-posterior stance posture. To solve the problem, it appears that some method of measuring segmental angulation will need to be used. Depth of spinal curves as measured in the MacEwen and Howe method fails to take into consideration real anatomical differences in the shape of the vertebrae which, in turn, affect spinal curvature. There are indications that variations in body build may affect the depth of spinal curves also. Heredity also probably plays a part in depth of spinal curvature.

Other problems to be solved have to do with finding easy, quick methods of indicating body landmarks and with finding what the critical landmarks are. The research on determining external body landmarks would need to be validated by roentgenographs.

Sampling. Another problem which must be met in posture tests is sampling. If the individual knows his posture is being evaluated at a specific time, the sample may not be closely related to his habitual posture and this is the position which is important. It is impractical to ask an individual to stand for a period of time awaiting evaluation. On the other hand, we have no tests of segmental alignment which can be applied while the individual is moving. There are many complications in working out such a test.

It seems likely that motion pictures would have to be used or at least a high-speed camera which could "stop" motion. Suggestions for using such equipment may be found in Chapter 5. However, equally important are such factors as using some method to indicate body landmarks which will permit movement and which will also permit the landmarks to be observed from various angles. Electric lights require wires which are cumbersome; adhesive tape tends to fall off and is not too easily seen; skin pencil markings are not usually clear in photographs. Perhaps the use of some of the newer colored masking tapes may be the answer. Another complication is that of using a skin marking to denote a deep skeletal landmark. Unless care is taken to avoid moving soft tissue over the skeletal landmark, these points may be very inaccurately located.

The arm tends to hide anything used to mark the hip joint. To solve this problem, it may be necessary to locate that joint by triangular methods, using landmarks which are observable on silhouettes.

Still another problem may be the establishment of a true vertical line of reference which will be available for comparing deviations. Such a line should be observable in all frames of a motion picture if that method is used.

Evaluating Dynamic Positions. It will be appreciated that, if it is difficult to measure stance posture satisfactorily, it is even more difficult, and certainly of greater importance, to measure the dynamic aspects of body mechanics. Many of the same problems present in measuring stance posture also exist in measuring dynamic posture.

In addition, there is the problem of the constant change in segmental relations of the body. What are the critical angles between body segments? Are these affected by factors such as flexibility and balance? If they are affected, how much should be attributed to normal and how much may be considered a deviation from normal? How can these critical angles be determined? What body landmarks should be used? What is the relationship between these angles?

For example, in stooping, what is the relationship between the trunk and the thigh? How does this angle at the hip compare with the angle between the thigh and the leg? Should the trunk be parallel to the lower leg? What landmarks can be used to establish the angles? All the above problems need to be investigated in this one activity.

Another problem is the selection of the critical tasks to be evaluated in the test. This seems to be of lesser importance if a method can be found to evaluate dynamic positions of the body. It seems probable that photography will have to be used in the early stages. But to be practical for use with large numbers, some nonphotographic method is necessary. This should be inexpensive and must not be time-consuming if it is feasible to use with large numbers.

OTHER BODY MECHANICS PROBLEMS

Although the measurement aspects of body mechanics loom as the most pressing problems, there are other problems also needing serious consideration.

Motivation. The problem of motivation is a serious one. An individual may have the best instruction available in body mechanics, but if he lacks the motivation to practice the skills the instruction may be fruitless. Few people have attempted such research. Tate's study (29) indirectly approached this problem by using motion pictures of students performing in a test situation for study by the group. Although the photographs were taken for evaluation purposes, the students using them improved more in body mechanics than students who did not have such a motivational device. References to psychological theories on motivation may be found elsewhere in this chapter. However, there are no studies available of direct application to body mechanics.

Motivational factors which warrant controlled investigation to see which are the most effective include an appeal to the ego, the effect of using frequent test situations, the use of competition between groups, the use of negative practice, and the motivation of fear. The latter has been used fairly successfully commercially by advertisers to sell their products. Such a motivational factor is not without some foundation in fact in body mechanics, for it has been stated by Shannon and Terhune (27) that poor posture with increased lumbar lordosis and compensatory kyphosis is the most common cause of chronic lumbosacral strain. It is also important to investigate at which age level each motivational factor has the most influence.

Validation of Postural Movements. The most efficient methods of lifting, pushing, pulling, stair climbing, and the like have been analyzed kinesiologically, but such analysis has not been experimentally validated in most cases. For example, authorities disagree as to whether one should kneel beside or directly behind an object to lift it. Should the housewife carry her basket of clothes in front of her or beside her on her hip? The list of body mechanics tasks to be studied is endless, and the only satisfactory way that answers to these problems can be obtained is by studying these tasks experimentally.

Stair climbing is another area of controversy. Should the whole foot be placed on the stair in ascending, or is it just as efficient to use the more natural way of placing only the ball of the foot on the step? This problem, as well as the other body mechanics problems, might be investigated on the basis of energy costs as measured by metabolic equipment or gas analysis, by measurement of the fatigue factor, or by electromyography. (See Chapter 6.) All three methods present difficulties. Energy costs and electromyographic studies require special equipment. It is extremely difficult to get a true measure of fatigue in studies of that type. The individual tends to slacken his efforts before true fatigue sets in. Strength measures tend to increase after fatigue bouts because of the warm-up effect, although balance measures do appear to be affected by fatigue.

Electromyographic studies of postural muscles in various movements have been done by Portnay and Morin (22). Their findings indicate that electrical activity of the muscles varies in different individuals. They also found that the muscles respond when stretched, because of the displacement of the center of gravity.

Energy costs of erect versus relaxed standing postures need additional investigation. Some early work in this area was done by Hellebrandt and her associates (6). McCormick's findings (17) indicated that erect posture necessitated greater expenditure of energy, but findings of British scientists (2) showed an increase of 30 to 50 percent in energy costs when the subject had to stoop to 80 percent of his height. While the average person does not stoop that amount when in a slumped position, there is enough of a question raised to warrant further investigation.

Relation to Socio-Economic Status and Health. The relationship of socio-economic status and of physical and mental health to posture is not known with any degree of certainty although some investigation has begun along these lines. Moriarity and Irwin (20) found that physical defects such as disease, fatigue, heart defects, underweight, and asthma occurred more frequently among children with poor posture. Emotional disturbances manifesting themselves as self-consciousness, a tendency to fidget, restlessness, and timidity were also more prevalent among children with poor posture.

The actual relationship of mild depression or anxiety states to posture needs documenting. Barlow (1) found that neurotic people tend to sway more than normal individuals because muscle tensions tend to interfere with their awareness of minor degrees of sway which are normal. Whether other factors—such as slumping or excessively rigid, extended postures—are associated with mental disturbances should be investigated.

Body Alignment. While body alignment varies normally from individual to individual and during various periods of growth, there needs to be a limit set beyond which certain body positions would not be considered normal. The establishment of norms for posture at various age levels would help in evaluating whether lumbar lordosis, for example, may be considered normal or abnormal for the age level under consideration. Some of the other areas where age norms would be helpful are in positions of the vertebral border of the scapula, the pelvis, and abdomen. The knee region needs to have standards set for the age at which it would be considered abnormal for knock knees or hyper-extended knees to remain. Longitudinal studies of changes in posture from that of the child to the adult might be used to establish norms for postural changes. It might be desirable to correlate postural change with skeletal age, established by roentgenographs of the wrist joint.

At the other end of the age scale, in view of increased interest in the aging population, it would be useful to know how much the erect position of the body normally deteriorates. In this same age range, a study needs to be made on the relationship of joint disabilities and various chronic diseases, such as arthritis, to poor posture. This might help provide motivational material for use with younger age groups.

Carry-Over Value. It has been assumed that there is a carry-over value of training in body mechanics. However, no studies investigating this premise have been published. Do the students who have been subjected to courses in body mechanics demonstrate better body mechanics and have less trouble with backaches from postural causes? Are they less susceptible to injury from faulty use of the body in lifting, carrying, and the like than their contemporaries who did not have such training? What are the needs of the various occupations that our students enter? Were they given adequate training in applying principles of good body mechanics in unfamiliar situations? Answers to these questions would assist in curriculum planning for courses in body mechanics.

Many Other Problems. There are other problems related to body mechanics for which answers are needed. For example, how much strength is necessary in the various muscle groups involved in the maintenance of satisfactory posture? How does strength vary with body build? How important is the developing of strength in antigravity muscles? Do students usually have enough strength? Is it a matter of motivation or kinesthetic teaching for correct position as much as the development of strength?

Many problems in body mechanics await the individual wishing to do research, and many of the problems are vital to our welfare.

PROBLEMS OF THE FEET

Like posture, evaluation of the foot and its dynamics is a difficult problem. Basically, many of the problems are the same. Although it is believed certain deviations, such as pronation, may have an effect on foot function, there is little objective evidence for the beliefs. Feet which appear to deviate markedly may be symptomless, while others which appear to be normal may cause considerable disability because of pain or fatigue.

It is now known that height of the arch is not a satisfactory method of evaluating foot function, for a well-muscled foot may have the arch area partially filled in by muscle bulk. Furthermore, the highly arched foot may not be functional from the standpoint of withstanding long periods of weight-bearing.

Causes of foot instability have been studied by Morton (21), Jones (14), Willis (30), Keith (15), Steindler (28), and Schwartz (26). Their theories on foot disability vary from insufficiency of the extrinsic foot muscles to lack of ligamentous support or to imbalance of the foot caused by shortness of the first metatarsal. The deflecting of the body weight in the tarsal region of the foot from the center of the talus through the calcaneus, which makes its contact with the ground to the outside of the center line, has been advanced as a cause for the rolling of the foot inward.

Studies of weight distribution in the foot while it was involved in walking have been carried out by Schwartz (24, 25), who maintains that muscular contraction alone cannot prevent pronation and foot strain. He found that the posterior tibial muscle does not reflexly contract to prevent pronation either in stance or in the stance phase of locomotion, although this muscle is commonly accepted as having that function.

A very careful study of the axes of the ankle and foot joints has been carried out by the British anatomist, Hicks (12). He verified his findings on recently amputated specimens by roentgenographs of living subjects. In another study, Hicks (13) states that the plantar aponeurosis is attached at the distal end through the plantar pads of the metatarsal-phalangeal joints to the proximal phalanges. As a result of this attachment, when the toes are forced into the extended position in standing on the toes or in the push-off phase of walking, the arch rises by the ligamentous mechanism mentioned above without the direct action of any muscles. This observation has import for planning remedial work for the foot.

With basic disagreement among authorities on causes of foot disabilities, it is apparent that there is room for much research. Little is known about the relation of strength of the feet and legs and the incidence of foot difficulties. One of the problems has been how to measure strength of the intrinsic foot muscles. Inevitably, the strength of the leg muscles becomes involved in such measurement. Strength may be measured by devices such as the tensiometer, although it is difficult to use this device for small muscle groups. Another measuring device which measures finer units is the force indicator.

However, the measuring device is not the only problem. It is difficult to isolate the small muscle groups so that they can be evaluated. Electromyography or Schwartz' device (24, 25) could also be used. A study needs to be undertaken to find the relative strength of various foot and leg muscles in the painful foot compared with the strength of these muscles in the normal foot. This would then serve as a scientific basis for devising exercises to strengthen the foot and leg muscles.

While most authorities accept pronation as a normal part of the development of the walking pattern in children, there is little agreement at just what age this should be considered abnormal. Pronated feet are relatively common in teen-agers and young adults who do not complain of foot disability. Does this defect predispose them to foot troubles at a later age? No one seems to have the answer to that problem. How common is pronation among the middle-aged? Is it associated with foot disturbances? What are the causes of pronation? Is it an imbalance in the joints of the foot itself, a muscle imbalance of the lower leg, or is it due to imbalance in the rotator muscles at the hip? Does exercise have a place in correction of this deviation and, if so, what muscle groups need strengthening?

There is also lack of agreement on when knock-knees should disappear and just what the developmental pattern is for use of the foot and leg muscles. Study of the literature will reveal other points of disagreement or uncertainty.

Fortunately, infectious diseases involving high fevers and prolonged convalescence are on the wane. We have been careful to avoid exertion under these conditions to avoid straining the heart and its valves. But the heart is a muscle. If it is weakened under such conditions, might not the other skeletal muscles be similarly weakened? A study needs to be made of the weakening effect of disease and convalescence on muscles involved in posture and the feet.

BIBLIOGRAPHY ON BODY MECHANICS

1. BARLOW, WILFRED. "Psychosomatic Problems in Postural Re-Education." *The Lancet.* September 24, 1955.
2. BEDFORD, THOMAS, and WARNER, C. G. "The Energy Expended in Walking in Stooping Postures." *British Journal of Industrial Medicine* 12:290-95; October 1955.
3. CURETON, THOMAS K., and WICKENS, J. STUART. "Center of Gravity in the Antero-Posterior Plane and Its Relation to Posture, Physical Fitness, and Athletic Ability." *Research Quarterly*, Supplement 6:93-105; May 1935.
4. ELFTMAN, HERBERT. "A Cinematic Study of the Distribution of Pressure in the Human Foot." *Anatomic Record* 59:481-91; July 25, 1934.
5. FOX, MARGARET G., and YOUNG, OLIVE G. "Placement of the Gravital Line in Antero-Posterior Standing Posture." *Research Quarterly* 25:277-85; October 1954.
6. HELLEBRANDT, F. A.; BROGDAN, ELIZABETH; and TEPPER, RUBYE H. "Posture and Its Cost." *American Journal of Physiology* 129:773-81; June 1940.
7. HELLEBRANDT, F. A., and FRIES, E. CORINNE. "Constancy of Oscillograph Stance Patterns." *Physiotherapy Review* 22:17-23; January-February 1942.
8. HELLEBRANDT, F. A., and KELSO, L. E. A. "Synchronizing Biplane Stance Photography with Center of Gravity Observations." *Physiotherapy Review* 22:83-87; March-April 1942.
9. HELLEBRANDT, F. A.; KELSO, L. E. A.; and FRIES, E. CORINNE. "Devices Useful to the Physiological Study of Posture." *Physiotherapy Review* 22:10-16; January-February 1942.
10. HELLEBRANDT, F. A.; RIDDLE, KATHRYN S.; and FRIES, E. CORINNE. "Influence of Postural Sway on Stance Photography." *Physiotherapy Review* 22:88-97; March-April 1942.

11. HELLEBRANDT, F. A. and OTHERS. "The Location of the Cardinal Anatomical Orientation Planes Passing Through the Center of Weight in Young Adult Women." *American Journal of Physiology* 121:465-70; February 1938.
12. HICKS, J. H. "The Mechanics of the Foot." *Journal of Anatomy* 87:345-57; October 1953.
13. HICKS, J. H. "The Mechanics of the Foot: II. Plantar Aponeurosis and the Arch." *Journal of Anatomy* 88:25-30; January 1954.
14. JONES, FREDERIC WOOD. *Structure and Function As Seen in the Foot.* Baltimore: Williams and Wilkins Co., 1941.
15. KEITH, ARTHUR. "History of the Human Foot and Its Bearing on Orthopedic Practice." *Journal of Bone and Joint Surgery* 11:10-32; January 1929.
16. KELLOGG, J. H. "Observations on the Relationships of Posture to Health and a New Method of Studying Posture and Development." *Bulletin of Battle Creek Sanitarium and Hospital Clinic* 22:192-216; September 1927.
17. MCCORMICK, HARRIET G. *The Metabolic Cost of Maintaining a Standing Position.* New York: Columbia University Press, 1942.
18. MACEWEN, CHARLOTTE G., and HOWE, EUGENE C. "An Objective Method of Grading Posture." *Research Quarterly* 3:144-57; October 1932.
19. MASSEY, WAYNE W. "A Critical Study of Objective Methods for Measuring Antero-Posterior Posture with a Simplified Technique." *Research Quarterly* 14:3-22; March 1943.
20. MORIARITY, MARY J., and IRWIN, LESLIE W. "A Study of the Relationship of Certain Physical and Emotional Factors to Habitual Poor Posture." *Research Quarterly* 23:221-25; May 1952.
21. MORTON, DUDLEY J. *The Human Foot.* New York: Columbia University Press, 1935.
22. PORTNAY, HAROLD, and MORIN, F. "Electromyographic Study of Postural Muscles in Various Positions and Movements." *American Journal of Physiology* 186:122-26; July 1956.
23. REYNOLDS, E. S., and LOVETT, ROBERT W. "Methods of Determining the Position of the Center of Gravity in Relation to Certain Bony Landmarks in the Erect Position." *American Journal of Physiology* 73:286-93; May 1909.
24. SCHWARTZ, R. PLATO; TRAUTMAN, OTTO; and HEATH, ARTHUR L. "Gait and Muscle Function Recorded by the Electrobasograph." *Journal of Bone and Joint Surgery* 18:445-54; April 1936.
25. SCHWARTZ, R. PLATO, and OTHERS. "Kinetics of Human Gait; Making and Interpretation of Electrobasographic Records of Gait." *Journal of Bone and Joint Surgery* 16:343-50; April 1934.
26. SCHWARTZ, R. PLATO, and OTHERS. "Useful Methods of Examination as Related to Cause and Treatment of Painful Feet." *Physiotherapy Review* 19:19-24; January-February 1939.
27. SHANNON, PAUL W., and TERHUNE, S. RALPH. "The Low Back Problem." *American Surgeon* 17:1106-112; December 1951.
28. STEINDLER, ARTHUR. *The Mechanics of Normal and Pathological Locomotion in Man.* Springfield, Ill.: Charles C. Thomas, 1935.
29. TATE, MARJORIE B. *The Comparison of Two Teaching Methods on Learning the Mechanics of Selected Body Movements.* Doctor's thesis. Iowa City: State University of Iowa, 1956.
30. WILLIS, THEODORE A. "Function of the Long Plantar Muscles." *Surgery, Gynecology and Obstetrics* 60:150-56; February 1935.

Kinesiology and Activity Analysis

C. ETTA WALTERS
JOHN COOPER
OLIVE YOUNG

Kinesiology, the science of bodily movement, deals with overt human movement produced by the skeletal musculature and the associated nervous system. Its object in physical education is to explain human motor performance, co-ordination, and skill. Undoubtedly, the primary contribution of the physical educator to the study of motor behavior will be descriptive analysis. However, this descriptive analysis must be interpreted in terms of such foundation sciences as genetics, anatomy, physiology, and physics. These will be discussed here only briefly.

SCIENTIFIC FOUNDATIONS

Genetics (Characteristic Behavior). Similarities throughout the evolutionary scale, rather than differences, appear to be the rule in the neurological development of overt behavior. The primitive sequence is muscle, motor fiber, and sensory fiber; and the response to proprioceptive stimulation, produced by changes in posture, precedes exteroceptive stimulation. Development is seen to progress in a cephalic-caudal and axial to distal direction; and evidence points to the fact that individuation of movement arises from a total pattern response.

Coghill's observations (20) on the Amblystoma, a relatively simple and typical vertebrate, have provided extensive knowledge of the development of overt behavior and have stimulated a wealth of experimentation on other forms, including fishes, reptiles, birds, and mammals. He studied both embryonic and post-natal behavior. By cinematography he recorded the first reflex and voluntary responses of the vertebrate and correlated them with the growth and development of the neuromuscular system. Hines (63) has studied similarly the development and regression of reflexes, postures, and progression in the pre- and post-natal monkey.

Minkowski (88) was one of the first to make a controlled study of the development of human movement by stimulating human fetuses. Hooker (65) and Windle (122) have more recently analyzed ontogenetic development by this method. Gesell's monumental studies (46) on infant behavior and developing patterns have furthered our knowledge of characteristic human behavior, and Wild (121), in the field of physical education, has studied the maturation of the overhand throw by cinematography.

The works of Weiss (118) and Sperry (106) have elucidated the role of the central and peripheral systems in the control of movement. By the method of recombination of patterns (crossed nerves and muscles) in lower vertebrates, they have shown that the central nervous system (CNS) is not so plastic as has been previously supposed and that one movement is not substituted for another by some sort of permanent switching in the central pattern. It involves the participation of the higher centers, which develop a new type of action.

Anatomy (Structure). In early studies, the action of muscles has been inferred from the origin and insertion of muscles as determined from cadavers. The limitations in relying on this method are that one assumes a muscular action without considering the effects of other muscles acting on the joint or in the movement. Duchenne (39) used the electrical stimulation method. By stimulating the muscle under examination, he would note the action produced on the joint or joints controlled by the muscle. Although this method has advantages over the anatomical one as deduced from cadavers, it does not tell us the influence of the other muscles which may enter into the movement. It also is a superimposed movement and does not consider the volitional factor. While we have discovered from these methods what a muscle can do, we are not told what it actually does do under ordinary circumstances or in various positions and under different conditions.

The functional anatomy of joints and ligaments must be analyzed by methods other than the traditional ones of inference from cadavers. Cinematographic methods have helped to define the action of different joints in movements. The part performed by ligaments in the stabilization of joints is being re-examined in light of new electromyographic evidence (18), which shows that the ligaments play a bigger part in maintaining posture than has been hitherto supposed.

The muscle itself is being recognized more and more for the complicated mechanism that it is, and the student of movement analysis must keep up with the literature on the chemistry of muscular contraction (111), blood supply to muscles, muscle fibers, and the intricate nerve supply and how it influences movement (81).

Physiology (Functioning). By means of microelectrodes inserted into muscle fibers and by contact with individual nerves, Adrian and Bronk (1) developed a method for the cellular analysis of nervous activity. This consisted essentially of recording the activity of a single neuron. Through the extension of such techniques, we are learning much about the sensory control of movement. The auto-genesis of muscular contraction and inhibition, the double motor innervation of the muscle spindle, and the integration of the supra-spinal centers with the spindle, and thus with movement, have all been learned through such methods (47, 82, 87).

Seyffarth (99) has shown that the same motor unit comes into action in the same sequence when an identical movement is voluntarily repeated many times. If the movement is varied slightly, the motor units also vary. He has pointed out that two movements performed simultaneously are not a simple summation of the movements, but involve a new synergy.

Bosma and Gellhorn (9) have stimulated the motor cortex of monkeys and have recorded the electromyograms of the muscles of the arm when it was placed in various positions and under different conditions. Extending the study to include humans, Gellhorn (45) has shown the importance of muscle stretch in increasing muscular response, the effect of proprioception on movement, and the patterned response which results from different degrees of stress and its dependence upon the type of movement performed. Hellebrandt and others (53) have also demonstrated the latter, and have shown the facilitatory influence of reflexes on work output.

Penfield (91) has stimulated, under local anesthesia, the cortex of conscious men and women, has recorded the ensuing movement, and, as a result of extensive observation, has written about the engrams of movement in the CNS.

By ablation experiments, in which various parts of the brain and lower centers were extirpated, movement has been analyzed in terms of the contribution of various parts of the CNS (100). A study of brain-damaged and pathological cases has afforded another means of attaining this objective (5).

By means of electrical stimulation of the bulbar reticular system and the recording of electromyograms of selected muscle groups, Magoun (85) has shown the importance of this part of the brain stem in integrating movement and has provided the rationale for understanding some of the reflex aspects of posture and locomotion. These experiments have also helped to explain the effect of emotions and excitement on muscular performance.

In 1895, Richet (96) demonstrated that the soccer kick was accomplished by a *contraction ballistique* in which the quadriceps impelled, rather than dragged, the limb; and the actual blow was effected by the momentum of the limb after the contraction had ceased. The term *ballistic* has been used by some authors to denote throwing and jumping movements, which are in a sense ballistic, in that the body or an object is projected and travels through space by momentum according to physical laws. But the term as it was used by Richet and subsequent workers refers not to the purpose of the movement but to the physiological conditions under which the movement is produced and to the nature of the movement itself. Although skilled throwing and jumping movements are ballistic, a ballistic movement is not a *throwing* movement, but a *thrown* movement in which the limb is impelled, rather than pulled, by the driving muscle and in which there is a momentum phase free of muscular action.

Electromyographic studies of fast and slow movements, and of movements produced with varying loads, show that the neuromuscular pattern varies according to the speed and the load (54, 107). Some of the earlier works demonstrating this came out of Stetson's laboratory at Oberlin. The results of these and similar studies have changed our concept of the relationship of agonists and antagonists in movement and have shown the fallacy of attempting to analyze movement by anatomical facts alone. They have also pointed out the wisdom of demonstrating and of analyzing movement in the tempo at which it is to be executed.

It is common knowledge, and has been verified by energy cost experiments, that it is easier to walk downstairs (negative work) than upstairs (positive work). In either case, the force opposing gravity is muscular tension, and the muscular tension necessary to maintain the body in motion at the same speed must be the same, whether the movement is uphill or downhill. Thus, for an explanation, we must look to muscle physiology. The length-tension diagrams, in which tension of isolated or whole muscles is recorded for the muscle at varying lengths, has been studied by Blix (8) and others (60, 92). This gives us some basis for judging the muscle's capacity for producing tension under different conditions. Blix found that tension increases when a muscle is stretched during an isometric tetanus to values well above those of the isometric maximum, and when a muscle shortens during tetanus its tension falls far below the tension of the isometric maximum. Thus we see that, if a muscle shortens during contraction, there must be an increasing number of active fibers brought in to maintain the tension of the whole muscle, while a muscle which is lengthening during its contraction must have an increasing tension built up in each active fiber, with a subsequent number of active fibers decreasing during movement. Although many have studied the effect of positive work, few have studied negative work. Chauveau in 1896 (16) was one of the first to do so. This is a relatively untouched area in kinesiology, and yet the implications for using a movement involving negative work, which at low and moderate velocities of shortening has been shown to be one-third to one-ninth that of positive work, wa[illegible] further investigation (4).

Physics (Mechanical [illegible]ciples). Human movement, as movement, involves principles of mechanics (3, 48, 66, 83), the branch of physics which is concerned with relations of mass, space, and time. (See Black, 7.) From a mechanical standpoint, human movement may be considered as some mass moving through some space at some rate, or as displacement with respect to time. Thus, much human movement and many athletic performances may be considered as essentially the application of internal, controllable forces to man's center of mass in order to overcome (temporarily) the resistance of constant external forces, such as gravity, and to produce some translation or rotation of the body. The Webers in 1836 (117) considered human locomotion as the result of simple pendulum movements of the limbs. Braune and Fischer in 1885-95 (11) showed that the force of gravity was insufficient to produce the acceleration of the limb observed in walking, and established the basis for computing the human center of gravity.

It is customary to divide mechanical analyses of human movement into two general classes—external and internal body mechanics. Analyses of man's effectiveness in overcoming external resistance in order to propel himself or some object are considered studies of external body mechanics.

Human movement also has an internal mechanical aspect, since anatomical relations are essentially mechanical and the tissues are structural material. Analyses have been made of the mechanical action and interaction of bodily segments. As examples of such studies of internal body mechanics involving movement, those done by Fenn (41), Elftman (40), and Steindler (108) could be cited.

METHODS AND DEVICES FOR MOVEMENT ANALYSIS

This is not meant to be a detailed list of all methods and devices used in the analysis of movement. It is a brief recapitulation of the methods, as reported in various studies, which the physical education graduate student and beginning research worker might employ.

Photography (26, 51). This is one of the most widely used methods in the study of movement. It includes cinematographic, stroboscopic, and X-ray analyses. The method is well covered in the section on Photography in Chapter 5, and the reader is referred to that section for a detailed description of equipment and method.

While cinematography may be the medium for obtaining a record of the activity, considerable study is necessary in the analysis obtained from the picture. There are various methods used in this analysis.

In 1925, A. V. Hill (61) pointed out that the ordinary laws of mechanics were applicable to all types of jumping, and since that time the majority of studies have made use of these laws (27, 28, 44, 72). Cureton (27), in particular, applied the physical laws of projection to high jumping and broad jumping and showed that any jump could be analyzed in terms of (a) initial take-off velocity, (b) angle of take-off, and (c) time in the air. These values, when substituted in the following equations checked very closely with the actual distance of the jump.

Height of Jump $= V_1 \sin \theta t - \frac{1}{2} gt^2$
(in feet)
Distance of Jump $= V_1 \cos \theta T$
(horizontally)

V_1 = ft./sec. initial take-off velocity
θ = angle of take-off, center of gravity to block
T = time of flight in sec.
t = ½ time of flight in secs.
g = 32.2 ft./sec.^{2}

Patterns of movement have been studied by photography. Wild's study (121) is an example of methods used in studying characteristics and temporal relationships and velocities of different segments of the body at various stages in the development of the overhand throw. She analyzed her film by measurement of actual distances traversed by the ball in a given period of time; by verbal description of the photographed movement; and by tracing positions of the body, arm, and hand at crucial stages of the throw. By putting all movement and timing items of each throw into the proper age list, she was able to show patterns of the various phases of the throw, the whole throw, and characteristics of ages and sex. Zimmerman (123) studied the movement of the center of gravity and the leg and arm movement patterns in skilled and nonskilled college women in the standing broad jump. The amount of flexion and extension of various joints, as well as the duration of the movements, was measured at various phases in the jump.

Time Analysis. The timing of reflexes, reaction time, and speed of movement is a method whereby movement may be analyzed. It is definitely used in cinematographic analysis and is necessary in almost all methods in which movement is recorded. For a description of timing devices, Chapter 6 on equipment should be read. Specific studies, especially those reported in the section below on research in fundamental activities, give methods using these devices. Bowne (10) constructed an ingenious timing device for her study in the underarm throw. It can be readily made at a low cost.

Movement Recording Systems. Attempts to record normal human movement have employed lever systems, which have not proved very satisfactory. Hubbard (68) has used thread and rubber band systems successfully. They produce time displacement or velocity curves without radial distortion. While the beginning research worker should be aware that there are such methods in studying movement, they are perhaps more appropriate for experienced investigators.

Muscle Action Recording Systems. Several methods and devices are used in recording muscle tension, strength, force, and endurance.

Muscle Tension. Muscle hardening (*deformation* or *bulge*) has been used as a basis for determining the presence or absence of muscular tension during certain periods of a movement cycle in several studies of normal human movement by Stetson and Bouman (109) and Hubbard and Stetson (69). There are two methods of picking up the evidence of muscle hardening, but both depend on transmitting this evidence by means of an air column enclosed in rubber tubing—pneumatic recording. Since the air column is at or near atmospheric pressure, an average correction of milliseconds for each foot of tubing is subtracted before relating the tracing to the movement tracing (13). The chief difference in the two methods is the type of applicator used.

One type of applicator is similar to a Marey tambour, but it is smaller and has no lever system. The applicator consists of a light cup with a condom rubber diaphragm to which a small cork boss is cemented. When the applicator is taped securely to the skin over the muscle, the elasticity of the rubber diaphragm forces the cork boss into the skin and deforms the underlying muscle slightly. As the muscle hardens, the boss is forced out and displaces the diaphragm (70).

The other type of applicator consists of a small cup with no diaphragm, which is held firmly to the skin over the muscle by the negative pressure (suction) of a siphon aspirator. The suction draws the skin slightly into the cup; when the muscle hardens, the skin under the applicator tends to flatten and this change is transmitted through the pneumatic system to a recording pneumodeik. The pneumodeik, in this case, must either be built to withstand negative pressure, or a septum must be placed between the aspirator and the pneumodeik.

The recording pneumodeik by Hudgins and Stetson (70) is an improvement of the familiar Marey tambour. It is designed for kymographic recording, but can be adapted for oscillographic recording. Brown (13) has discussed some of the technical aspects of pneumatic recording systems.

Muscle Activity. Action current recording devices—electromyography (6, 35, 45, 115, 53, 54)—record muscle activity. An electromyogram is a recording of the electrical activity in a muscle. Since a normal muscle displays electrical silence at rest, the presence of electrical activity in a muscle denotes activity. Surface electrodes are more suitable for measuring activity in kinesiological studies than needle electrodes. The latter produce insertion activity and cause pain upon movement, thus introducing a variable factor. Action potentials of the muscles can be recorded on an oscillograph or on paper by an ink-writing dynograph or crystograph. The latter is probably the most useful and widely used recording method in movement analysis.

Electromyography cannot be used to determine the tension of inaccessible muscles, and one must view with caution the results when proper procedure and analysis have not been made in accordance with the limitations of the method. It does afford a means of studying temporal relationships of muscles and agonist-antagonist relationships. With improved techniques brought about by better equipment, it will probably be one of the most valuable aids in the study of movement analysis. Like all other methods of kinesiological analysis, it is improved by using other methods in combination with it. Hellebrandt and others (53) have used it in combination with photography and ergography. Electromyograms were recorded as the individual performed bouts of wrist flexion or extension on a wrist ergograph. The work output was recorded objectively from the ergograph, and correlated with the integrated sine wave microvolt output of a selected muscle. Serial 35mm photographs were taken of the subject and subjected to analysis in a microfilm reader. Analysis of the photographs with the electromyograms and the ergograms gave data necessary for a more adequate analysis than would be possible with electromyograms alone.

Chapter 6 on equipment gives a more detailed description of equipment needed in electromyographic analysis.

Muscle Strength, Force, and Endurance. Dynamometers, strain gauges, tensiometers, ergographs, and similar instruments have been used to measure strength and endurance. While much research in this area belongs in the section on physiology, the factors affecting movement have been analyzed by instruments and devices used to measure strength. Clarke (17) has used the tensiometer to measure strength of selected muscle groups before and after participation in certain activities, and from this he has determined the muscles used most in the activity. Hellebrandt (52) has used the ergograph to study the effect of alternating and reciprocal movements on work output and thus on movement efficiency. For a description of the use of force plates in studying movement, the reference by Rehman and others (95) is suggested.

Globographic Technique. Dempster (37) describes a method based on the Albert-Strasser globographic techniques of measuring joint movements in cadavers and suggests that this method might be used for study on joint analysis in the living.

Comparative Anthropometry. (See section above on anthropometry.) One can, by statistical analysis of physical characteristics of highly skilled and average performers, determine the physical characteristics which are conducive to superior performance. Krakower (80) used this method in assessing anthropometric measurements influencing success in the high jump. He correlated the length of the legs, height, and breadth of the foot with high jumping. The length of the leg was considered in proportion to the length of the trunk. By a comparison of such relationships, he was able to show that every good jumper was shorter in height than would be expected from the length of legs and breadth of foot.

Watson (116) studied the relationship between throwing ability and certain body measurements in college women. By means of individual and multiple correlations, she demonstrated that there was a very low relationship between the ability to throw a baseball and body measurements in the subjects of her study.

Energy Cost. (See the section below on physiology.) The efficiency of a movement can be measured by the oxygen cost of the activity (23). While certain playing situations do not lend themselves to this method, it has been used in a variety of activities (15, 62, 76).

Asmussen (4) has studied the energy cost in positive and negative work by investigating bicycling uphill and downhill on a motor-driven treadmill. By using this method, the cost of the work can be measured by means of a Douglas bag and the work may be varied by changing the slope or speed of the treadmill. The conditions under which the muscles work can be varied by changing the rate of pedaling or by changing the length of the pedals and by varying the height of the bicycle seat. This affords a method of controlling some of the variables which are present in the investigation of treadmill walking.

RESEARCH IN FUNDAMENTAL ACTIVITIES

Mechanics of Starting in Track. Experimental studies on track starting represent a special type of research, the result of which may be applied to the initiation of movement in many activities. Most of the early studies were conducted during the 1930's. Several types of timing devices were used by these early investigators—the chronograph, chronoscope, electrically operated stop watches, and especially devised photoelectric beams and switches.

Some chrono-photographic devices have been used in top collegiate and Amateur Athletic Union track meets. However, they have been used for the purpose of recording the exact time for running the distance and not necessarily for research purposes.

Probably the most practical method of timing is by cinematography. Most cameras are either spring driven or electrically driven. A spring-driven camera does not run at a constant speed, owing to the decrease in tension on the spring as it unwinds. However, time may be recorded on the film by means of an electric clock, a synchronous motor device, a tuning fork, or falling objects. The electrically driven camera runs at a constant speed, and time may be computed by the speed of the cycle.

A phonograph motor is an example of a synchronous motor that may be used as the basis for constructing a timing device that may be used in cinematography. An especially constructed dial with a pointer or hand may be used to determine time. The angles the pointer makes must be computed in order to determine time accurately.

A camera with a speed of 64 frames or more per second produces the best pictures for purposes of analysis. Reasonable accuracy may be obtained by merely counting the frames and computing the average time for each frame. The flash of the gun or the movement of a starter's hand downward, compared with the first movement of the runner as shown on the film, makes it possible to determine the reaction time of the runner.

A Recordak film reader may be used to study the results of action recorded on film. This is a device that enlarges the image many times its original size. It may be run at a very slow or fast speed. The investigator may trace, record, or make various measurements as he views the film.

Most of the early investigators used improvised electrically devised starting blocks and various kinds of timing circuits in securing their results. Some examples of the areas they investigated are included here. Nakamura (90) and Walker and Hayden (114) attempted to find out the optimum length of time a sprinter should be held at the "get set" position. Westerlund and Tuttle (119) investigated the relationship between reaction time and speed in sprinting short distances. Dickinson (38) studied foot spacing at the start in relation to sprinting. Kistler (78) made a study of the distribution of force exerted upon the blocks from various foot positions. Others such as White (120) studied the effect of hip elevation on starting time.

The most recent studies (19, 24, 50, 55, 56, 57, 58) corroborate some of the earlier findings, contradict others, and bring to light new evidence. For example, foot spacings and foot arrangement at the start (24, 50, 55) were reinvestigated, as well as reaction time in reference to speed in running (25, 58). Also, the old theory that the "motor set" concept produces the fastest start was disproved by Henry (56).

The recent studies by Henry (55, 56, 58) made use of a chronograph with starting blocks constructed with calibrated springs attached to recording pens. The apparatus was designed to study time-force characteristics of the runner at the start. Cooper and others (24, 50) used the same chronograph and later used specially designed hydraulic starting blocks that measured force. The blocks were connected with an electrical starting and stopping system measuring time by a clock in one

one-hundredths of a second. Both force and time were recorded through the use of a special Bell and Howell camera designed to take pictures at the rate of 128 frames per second.

Mechanics of Running. Running is involved in almost every sport in which man participates, but there is some difference in the manner and style of running for track and for football. However, when fast sprint running or slow endurance running is desired, the form of running is essentially the same for every sport or activity.

The mechanics of running have been of interest even before Homer wrote of Achilles and his quickness of foot. Father taught son in the early dawn of history how to run for escape or capture.

In 1836 the Weber brothers (117) in Gottingen established with crude methods that the faster a man runs, the longer his stride becomes. Marey (86) and Demeny (36) proved this in France by using an improved photographic analysis.

Hill and Lupton (62) measured the speed of runners at intervals along the track, computed acceleration from the time records, and demonstrated the relationship between oxygen required and speed. On the basis of oxygen used, they were able to compute the foot pounds of work done and the mechanical efficiency of runners. Hill (59) applied the locomotion formula (Mass times Acceleration = Propelling Force — Resistance) to illustrate that the viscous resistance of the muscles and joints is the principal factor limiting the speed of the limbs in top-speed sprinting. He later modified his position concerning the importance of muscle viscosity.

Fenn (41, 42, 43) performed several studies on the center of gravity, work, kinetics, and friction in running. His cinematographic analysis at 120 frames per second is one of the most scientific of the recent studies. In this he demonstrated that a good sprinter developed about 13 horsepower, of which some 5.2 could be attributed to the initial energy and 7.8 to waste in recovery. Only 2.95 horsepower could be attributed to useful work directly related to propulsion, equivalent to 22.7 percent efficiency as a ratio of useful to total work. Wasteful work included work against gravity, wind resistance, ground contact resistance, and recovery energy associated with nonpropulsive limb movements. Fenn used a telephoto lens to reduce error of measurement, with the runner photographed against a background of lattice having squares one meter in size. Fischer's method was used to compute the center of gravity location progressively throughout the cycle. Fenn constructed a fast contact platform to show the component of retarding force graphically. He attributed the difference of .16 horsepower between contact and propelling energy to wind resistance.

Cureton (34) diagrammed the various styles in vector analyses and explained the action of the muscles in running based upon a review of various studies made by him and his students.

In a study of wind resistance (33) with a direct recording model of a man, mounted on the front of an automobile, a velocity versus air-resistance curve was developed. The effect of air-resistance was shown to approximate 2.8 percent of the propelling force at 30 ft./sec. An average sprinter running against a headwind would be slowed .57 sec. The power was calculated as .195 horsepower to overcome air-resistance on a quiet day.

Hubbard (67, 68) demonstrated in a combined movement and muscle action study that the "ballistic throw" of the limbs was used by trained runners and, as a result, they ran relatively more relaxed than the untrained runners.

The fact that the legs can alternate in cycling at a rate considerably faster than the fastest attained in sprinting, as confirmed by Slater-Hammel (102), indicates that the speed of the neuromuscular mechanism is not the factor that limits leg speed in sprinting.

Running may be studied in many ways. The treadmill may be used to record work and, as a result, endurance. The energy output can be ascertained by many methods that are in the section on physiology. The results of such studies, for example, reveal the relationship between energy and speed in horizontal running.

Some of the same pieces of research equipment previously mentioned as being used in track starting have also been used in studying running. Henry (58) attached a thin wire from a roll to the runner and, with the use of his specially designed chronograph, was able to determine the runner's speed at any given distance and time from the start to the finish of a race. Electrically wired gates connected with a clock and a recording device enable the investigator to determine velocity, acceleration, and deceleration in relation to the distance traveled. The same information may be secured by using specially devised photoelectric beams and switches. Force plates such as used by Rehman, Patek, and Gregson (95) in studying gait could be used to record pressures exerted by the foot against the ground in running.

The method of taking X-ray motion pictures of walking (94) may also be used for studying running. The problem would be to construct a large enough screen and running track to be able to record the movements accurately.

The length of stride may be easily measured by marking off certain standard distances on the ground. It is easy to plot the footprint distances. Covering the runner's feet with chalk enables the experimenter to determine the length of the stride more easily. Collins (21) measured the stride of the runner on the treadmill by placing painted bands known distances on the tread. Hogberg (64) studied length of stride, stride frequency, flight period, and maximum distance between the feet during running with different speeds.

Cinematographic analysis of the movement of the center of gravity will be enhanced by making special markings on the subject's hip and using a grid background such as Fenn used (43).

The recent research on warm-up by Karpovich and Hale (75) raised some questions concerning the long accepted value of a warming-up prior to running. More investigation of the topic with more subjects and in many movement experiences is needed.

Physical education students should continue to be challenged by the possibilities of attempting to find out why man moves as he does in speed events, and how he may improve his methods of overcoming his own inertia. A knowledge of track running, adequate subjects, proper instrumentation, and better treatment of data will help in the conduct of such experiments. Even if no new methods are discovered, the substantiation of previous findings alone is justification for periodic research being conducted in this area.

Analysis of Human Locomotion (Walking). The first studies of human locomotion were morphological descriptions of the gait.

An excellent summary and synthesis of these early studies which are basic to modern concepts of the mechanical analysis of human locomotion are found in the section on the gait in Steindler's scholarly work (108) on kinesiology. Steindler presents a detailed morphological description and graphic presentation of the gait, as well as analyses of various types of pathological gaits.

The earliest aid to the objective analysis of the gait was photography. By means of cinematography, it is possible to record the sequence, duration, and synchronization of each phase of the walk or gait. These exact recordings of the movement patterns in locomotion make it possible to determine extensive and detailed measurements. For example, Huelster (71) used photography as the basis for determining the reliability of lateral deviation during the supporting phase of the walk. Bilateral contour asymmetry measurements were made on uniformly enlarged photographs from selected motion picture frames. Slocum (105) used stroboscopic photography to analyze the gaits of various types of leg and foot amputees.

Morton and Fuller (89) used motion picture photography combined with barographic studies to analyze movements of the center of gravity of the body with relation to supporting contacts of the feet. The barograph makes it possible to measure the length of time the foot is in contact with the ground. Contact indicators show by electric lights when the heel or forepart of the foot is in contact with the ground and when this contact or pressure is released.

It is also possible by use of the barograph or electrobarograph (98) to study pressure forces involved in the walk. By means of carbon discs with isometric sensitiveness, it is possible to make direct measurements from pressure curve diagrams of the distribution of gravitational reaction from the floor in normal and pathological gaits.

Myokinetic studies of the gait have been concerned with the establishment of patterns of sequential muscle action of the lower extremities. The first studies used gross palpatory methods to determine which muscles were contracting. More recently, electromyography has made the analysis of contraction patterns and sequential action during the gait more precise.

Extensive research conducted under the direction of the National Research Council's Committee on Artificial Limbs (79) has developed various methods of analyzing both normal and pathological gaits. One phase of this study was designed to obtain data on locomotor patterns of both the normal subject and the amputee. These particular studies were concerned with:

1. Displacement-time data obtained by the use of an interrupted light technique and used to determine the velocity and acceleration of any point on the leg
2. Analysis and comparison of various gait patterns using cinematography and a newly developed glass walkway
3. A study of foot pressure patterns using a barograph in connection with the glass walkway
4. Evaluation of the gait of the amputee by means of high speed motion pictures
5. Measurement of forces acting upon the leg during locomotion by means of force plates which measured vertical load, fore-and aft shear, lateral shear, and torque
6. Analysis of the rotation occurring during locomotion by means of photographic records of movements of targets attached to specified points on the legs.

RESEARCH IN MECHANICS OF SPORTS ACTIVITIES

The research worker who attempts to analyze skillful performance in sports is faced with both the problem of the neuromuscular responses and co-ordinations and the problem of understanding and correctly applying certain mechanical principles. The analysis of motor performance may be approached from either of these aspects, or from both.

Swimming. Much of the research in swimming has been concerned with the propulsive forces which cause the body to move through the water, the resistance to this progression, and the buoyancy of a body in the water.

Alley (2) has classified studies of resistance and propulsion in swimming into four general categories. First, there are those studies which measure "drag" or the resistance offered to a body as it is towed through the water at varying speeds (74). The second group contains those studies in which the propulsive force of the swimmer is measured. In this procedure the velocity is zero, since the measuring device—an arrangement of ropes and pulleys attached to weights, spring scales, or dynamometers—prevents the swimmer from actually making progress (27). The third category includes studies of maximum speed attained by the swimmer as related to resistance and propulsion (73). The last group of studies includes those which are primarily theoretical in nature and which make use of formulas from classical hydromechanics (76). Included in this latter group were those early studies that first used vector diagrams to analyze water resistance into a supporting component of force and an impeding component of force.

The construction and development of various devices to measure resistance and propulsive forces have made possible definite advances in the scientific analysis of swimming mechanics. Alley (2) and Counsilman (25) developed and used similar towing devices to investigate the problems of water resistance and effective propulsive force. In the first study, the towing apparatus was constructed so that it controlled the velocity of the swimmer attached to it and simultaneously recorded on a kymograph the force exerted by the swimmer as he swam away from the apparatus or was towed toward it. From the data gathered it was possible to analyze the effective-propulsive force of various types of strokes. The towing apparatus used by Counsilman was designed to allow the attached swimmer to be towed or released at ten controlled velocities. It was possible to measure and record the drag, the effective propulsive force, and the fluctuation in propulsive force of the crawl stroke at each of the controlled velocities.

Studies on swimming starts have followed much the same patterns as those used in studying starting in track. In both cases the investigators have used cinematography and mechanical methods of analysis extensively. Cureton (31) used electrical starting blocks to determine which type of start was most satisfactory in speed and distance of dive. Tuttle, Morehouse, and Armbruster (112, 113) found that the starting blocks were a disadvantage to well-trained swimmers. These studies and others have also attempted to establish the optimum holding time between the "set" and "go" signals.

Research on the specific gravity and buoyancy of the human body has helped to contribute to the understanding of how and why the body moves in water as it does. Many of the basic principles of specific gravity and buoyancy have been applied to modern floating and swimming by Cureton (30). A study of the floating ability of college women by Rork and Hellebrandt

(97) included tests for specific gravity, buoyancy, and equilibrium in the water.

Gymnastics, Apparatus, Tumbling, Diving. Most of the research in the area of gymnastics and tumbling has been concerned with mechanical analyses of the various stunts. An early German book on gymnastic stunts and sports included a discussion of principles of inertia, gravitation, motion, force, center of gravity, equilibrium, centrifugal force, leverage, wheels and pulley, shafts and revolutions, eccentric push, friction, and resistance as applied to gymnastics or sports.

In his laboratory manual, Cureton (32) outlined procedures by which students in physical education could make application of mechanical principles to gymnastics and other sports. The deductive application of mechanical principles to gymnastic activities has been made by McCloy (84) and has led many students of gymnastics to make objective studies of these stunts and activities. Bunn (14) in his book on the application of scientific principles to coaching has made mechanical analyses of certain gymnastic exercises, as well as various sport activities.

Cinematography seems to be the most satisfactory method for studying and analyzing performance in gymnastics, apparatus, tumbling, and diving (49).

Sports and Other Activities. It is exceedingly difficult, if not impossible, to analyze sport skills during the actual playing situation. To make an analysis, the researcher must isolate the skill from the game, either in the laboratory or at least under laboratory controlled conditions.

Cinematography and stroboscopic photography have been used in studies of sport skills as well as in research on track events, diving, and gymnastics. Through stroboscopic photography Rehling (93) analyzed the golf drive of a number of expert golfers to determine what factors such experts had in common. He also calculated the velocity of the ball from these pictures.

Breen (12) used cinematography to analyze the pitching form of a number of major league pitchers. In this study the investigator was able to determine and compare the angle of the arm at the time of the release of the ball, the length of the stride, the angle of the pitching arm, the optimum height of the lead leg, and the angle of the leg at the time the ball was released. Bowne

(10) described the torque action of the principal joints in the overarm and underarm throwing patterns of high school girls. She points out the importance of trunk rotation in the transverse plane. This torque action of the principal joints in batting has been described by Conrad (22).

Electromyography has been used increasingly to determine the muscle patterns of a variety of sport movements. Slater-Hammel (103, 104) has analyzed the golf swing and tennis drive electromyographically. In both these studies, action current technique was used and recordings were made of the arm movements, the moment of contact with the ball, and the action currents of the muscular contractions. The path of the movement was recorded by means of a light glass thread and rubber band system. A photoelectric system was used to record the moment of contact with the ball. The data from all three phases were recorded simultaneously on a Teledeltos Polygraph.

In a subsequent electromyographical study of the golf drive, Karr (77) obtained high-speed, multiple, and simultaneous recordings of muscle action, acceleration, and movement patterns.

Walters and Partridge (115) used an eight-channel ink writing Offner crystograph to make simultaneous recordings of abdominal muscles in their comprehensive electromyographical study of abdominal muscle function. Paired skin electrodes were used to make detailed and extensive observations on the abdominal muscle actions during a number of exercises commonly listed for strengthening the abdominals.

In a study by Sigerseth and McCloy (101), the functions of six muscles in movements of the upper arm at the scapulohumeral joint were investigated. The action potentials of the selected muscles were recorded by a four-channel Grass electroencephalograph equipped with four ink-writing recorders. Paired electrodes were used to record the action currents of each of the muscles being studied.

BIBLIOGRAPHY ON KINESIOLOGY AND ACTIVITY ANALYSIS

1. ADRIAN, E. D., and BRONK, D. W. "The Frequency of Discharge in Reflex and Voluntary Contractions." *Journal of Physiology* 67:119-51; 1929.
2. ALLEY, LOUIS E. "An Analysis of the Water Resistance and Propulsion in Swimming the Crawl Stroke." *Research Quarterly* 23:253-70; October 1952.
3. AMAR, JULES. *The Human Motor.* New York: E. P. Dutton & Co., 1920. 470 p.
4. ASMUSSEN, E. "Positive and Negative Muscular Work." *Acta Physiologica Scandinavica* 28:364-82; 1953.

5. BEEVOR, CHARLES. *The Croonian Lectures on Muscular Movements. Remarks on Paralysis of the Movements of the Trunk in Hemiplegia.* New York: The Macmillan Co., 1901.
6. BIERMAN, W., and YOMSHON, L. J. "Electromyography in Kinesiology Evaluations." *Archives of Physical Medicine* 26:206; 1948.
7. BLACK, N. H. *An Introduction Course in College Physics.* Revised edition. New York: The Macmillan Co., 1941.
8. BLIX, M. "Die Lange und die Spannung des Muskels." *Skandinavische Archiv für Physiologie* 5:173-206; 1895.
9. BOSMA, J. F., and GELLHORN, E. "The Organization of the Motor Cortex of the Monkey Based on Electromyographic Studies." *Brain* 70:127-44; 1947.
10. BOWNE, MARY E. *The Contribution of Selected Body Levers to Ball Velocity in the Underarm and Overarm Throws.* Unpublished Doctor's thesis. Madison: University of Wisconsin, 1956.
11. BRAUNE, W., and FISCHER, O. "Die bei der Untersuchung von Gelenkbewegungen anzuwendende Methode, erlautert am Gelankmechanismus des Vorderarms beim Menschen." *Abhandlungen der Mathematische-Physichen Klasse der Sachsischen Akademie der Wissenschaften,* (Leipzig) 21:153-322; 1895.
12. BREEN, JAMES. *Cinematographic and Mechanical Analysis of Major League Baseball Pitchers.* Unpublished Master's thesis. Urbana: University of Illinois, 1949.
13. BROWN, P. F. "Characteristics of Pneumatic Recording Systems." *Journal of General Psychology* 20:505-509; 1939.
14. BUNN, JOHN W. *Scientific Principles of Coaching.* New York: Prentice-Hall, 1955.
15. CATHCART, E. P.; LOTHIAN, N. V.; and GREENWOOD. "Note on the Rate of Marching and Expenditure of Energy in Man." *Royal Army Medical Corps Journal* 34:297; 1920.
16. CHAUVEAU, A. As reported in *Symposium on Fatigue.* (Edited by W. F. Floyd and A. T. Welford.) London: H. K. Lewis & Co., 1953.
17. CLARKE, H. H.; MATHEWS, D. K.; and SHAY, C. T. "Strength Decrement from Carrying Various Army Packs on Military Marches." *Research Quarterly* 26:253-65; October 1955.
18. CLEMMENSEN, SVEND. "Some Studies on Muscle Tone." *Proceedings of the Royal Society of Medicine* 44:637-46; 1951.
19. COGHILL, G. D. *Anatomy and the Problem of Behavior.* New York: Macmillan; Cambridge, England: University Press, 1929.
20. COGHILL, G. D. "Flexion Spasm and Mass Reflexes in Relation to Ontogenetic Development of Behavior." *Journal of Comparative Neurology* 79:463-86; 1943.
21. COLLINS, MURRAY ROLAND. *The Effect of Stride Length on Efficiency in Distance Running.* Unpublished Master's thesis. Los Angeles: University of Southern California, 1951.
22. CONRAD, RUTH. *A Mechanical Analysis of the Skill of Batting.* Unpublished Master's thesis. Madison: University of Wisconsin, 1951.
23. CONSOLAZIO, C. F.; JOHNSON, R. E.; and MAREK, EVELYN. *Metabolic Methods.* St. Louis: The C. V. Mosby Co.; 1951.
24. COOPER, JOHN M., and OTHERS. Chapter on sprints in Mortensen, Jesse, and Cooper, John M., *Track and Field.* New York: Prentice-Hall, 1959.
25. COUNSILMAN, J. E. "Forces in Swimming Two Types of Crawl Stroke." *Research Quarterly* 26:127-39; May 1955.
26. CURETON, T. K. "Elementary Principles and Techniques of Cinematographic Analysis." *Research Quarterly* 10:3-24; May 1939.
27. CURETON, T. K. "Mechanics of the High Jump." *Scholastic Coach* 4:9-12; 1935.
28. CURETON, T. K. "Mechanics of the Shot Put." *Scholastic Coach* 4:7-10, 1935.
29. CURETON, T. K. "Mechanics and Kinesiology of Swimming the Crawl Flutter Kick." *Research Quarterly* 1:87-121; December 1930.

30. CURETON, T. K. "Natural and Artificial Buoyance, Flotation, and Body Glance in the Water." *Beach and Pool* 7:272; 1933.
31. CURETON, T. K. *Objective Tests of Swimming.* Unpublished Master's thesis. Springfield, Mass.: International YMCA College.
32. CURETON, T. K. *Physics Applied to Health and Physical Education.* Springfield, Mass.: International YMCA College, 1936.
33. CURETON, T. K. "The Mechanics of Track Running." (Edited by T. R. Brining.) *Athletic Journal* 10:32-34; 1931.
34. CURETON, T. K. "The Mechanics of Track Running." *Scholastic Coach* 4:7-10; 1935.
35. DAVIS, R. C. "Methods of Measuring Muscular Tension." *Psychological Bulletin* 39:329-46; 1942.
36. DEMENY, GEORGES. *Mechanisme et éducation des mouvements.* Paris: F. Alcan, 1924. p. 249-347.
37. DEMPSTER, W. T. "The Range of Motion of Cadaver Joints: The Lower Limb." *University of Michigan Medical Bulletin* 22:364-79; 1956.
38. DICKINSON, A. D. "The Effect of Foot Spacings on the Starting Time and Speed in Sprinting and the Relationship of Physical Measurements to Foot Spacings." *Research Quarterly Supplement* 5:12-19; 1934.
39. DUCHENNE, G. E. *Physiology of Motion.* Translated and edited by E. B. Kaplan. Philadelphia: J. B. Lippincott & Co., 1949.
40. ELFTMAN, H. "The Action of Muscles in the Body." *Biological Symposium* 3:191-210; 1941.
41. FENN, W. O. "Frictional and Kinetic Factors in the Work of Sprint Running." *American Journal of Physiology* 92:583-611; 1930.
42. FENN, W. O. "Work Against Gravity and Work Due to Velocity Changes in Running." *American Journal of Physiology* 92:433-62; 1930.
43. FENN, W. O. "A Cinematographic Study of Sprinters." *Scientific Monthly* 32:346-54; 1931.
44. GANSLEN, R. V. "Mechanics of the Pole Vault." *Scholastic Coach* 17:7:24-29, 32; 1947; 17:8:1 18, 60; 1947; 17:9:52-57, 63, 64; 1947.
45. GELLHORN, ERNST. "Patterns of Muscular Activity in Man." *Archives of Physical Medicine* 28:568-74; 1947.
46. GESELL, A. T. *Studies in Child Development.* New York: Harper & Brothers, 1948.
47. GRANIT, RAGNAR. "Reflex Self-Regulation of Muscle Contraction and Auto-Kinetic Inhibitions." *Journal of Neurophysiology* 13:351-72; 1950.
48. GRAY, JAMES. *How Animals Move.* Cambridge, England: University Press, 1953.
49. GROVES, W. H. "Mechanical Analysis of Diving." *Research Quarterly* 21:132-44; May 1950.
50. HAGER, ROBERT. *An Experimental Study on the Effects of Various Foot Spacings in the Sprint Start.* Unpublished Master's thesis. Los Angeles: University of Southern California, 1957.
51. HANSON, THIDA OAKES. *A Method of Analyzing Human Body Movement from Motion Picture Film to Determine the Range of Body Movement in Any Plane of Motion.* Unpublished Master's thesis. Madison: University of Wisconsin, 1950.
52. HELLEBRANDT, F. A.; HOUTZ, SARA JANE; and EUBANK, R. N. "Influence of Alternate and Reciprocal Exercise on Work Capacity." *Archives of Physical Medicine* 32:766-76; 1951.
53. HELLEBRANDT, F. A.; HOUTZ, SARA JANE; PARTRIDGE, MARIAM J.; and WALTERS, C. ETTA. "Tonic Neck Reflexes in Exercises of Stress in Man." *American Journal of Physical Medicine,* 35:144-59; 1956.
54. HELLEBRANDT, F. A.; HOUTZ, SARA JANE; PARTRIDGE, MARIAM J.; and WALTERS, C. ETTA. "The Chandler Table: Analysis of Its Rationale in the Mobilization of the Shoulder Joint." *Physical Therapy Review* 35:545-56; 1955.

55. HENRY, F. M. "Force-Time Characteristics of the Sprint Start." *Research Quarterly* 23:301-18; October 1952.
56. HENRY, F. M. "Research of Sprint Running." *Athletic Journal* 32:30-33; 1952.
57. HENRY, F. M. "Track Men Should Abandon Old Ideas." *Science News Letter.* August 17, 1957.
58. HENRY, F. M., and TRAFTON, I. R. "The Velocity Curve of Sprint Running." *Research Quarterly* 22:409-21; December 1951.
59. HILL, A. V. *Muscular Movement in Man.* New York: McGraw-Hill Book Co., 1927.
60. HILL, A. V. "The Mechanics of Voluntary Muscle." *Lancet* 2:947-51; 1951.
61. HILL, A. V. "The Physiological Basis of Athletic Records." *Scientific Monthly* 21:409-28; 1925.
62. HILL, A. V., and LUPTON, H. "The Oxygen Consumption During Running." *Journal of Physiology* 56:32; 1922
63. HINES, MARION. "The Development and Regression of Reflexes, Postures, and Progression in the Young Macque." *Contributions to Embryology.* No. 196 (Carnegie Institute of Washington Publications) 541:153-209; 1942.
64. HOGBERG, PAUL. "Length of Stride, Stride Frequency, Flight Period, and Maximum Distance Between the Feet During Running with Different Speeds." *Arbeitsphysiologie* 14:431-36; 1952.
65. HOOKER, DAVENPORT. *The Prenatal Origin of Behavior.* Lawrence: University of Kansas Press, 1952.
66. HOWELL, A. BRAZIER. *Speed in Animals.* Chicago: University of Chicago Press, 1944.
67. HUBBARD, A. W. "Experimental Analysis of Running and of Certain Fundamental Differences Between Trained and Untrained Runners." *Research Quarterly* 10:28-38; October 1939.
68. HUBBARD, A. W. "Muscular Force in Reciprocal Movements." *Journal of General Psychology* 20:315-25; 1939.
69. HUBBARD, A. W., and STETSON, R. H. An Experimental Analysis of Human Locomotion." *American Journal of Physiology* 124:300-13; 1938.
70. HUDGINS, C. V., and STETSON, R. H. "A Unit for Kymograph Recording." *Science* 76:59-60; 1932.
71. HUELSTER, LAURA J. "The Relationship Between Bilateral Contour Asymmetry in the Human Body in Standing and Walking." *Research Quarterly* 24:44-55; March 1953.
72. JOHNSON, BARBARA. *An Analysis of the Mechanics of the Take-off in the Standing Broad Jump.* Unpublished Master's thesis. Madison: University of Wisconsin, 1957.
73. KARPOVICH, P. V. "Analysis of the Propelling Force in the Crawl Stroke." *Research Quarterly* 6:49-58; May 1935.
74. KARPOVICH, P. V. "Water Resistance in Swimming." *Research Quarterly* 4:21-38; October 1933.
75 KARPOVICH, A. V., and HALE, CREIGHTON J. "Effect of Warming Up Upon Performance." *Journal of the American Medical Association* 162:1956.
76. KARPOVICH, P. V., and PESTRECOV, K. "Mechanical Work and Efficiency in Swimming Crawl and Back Strokes." *Arbeitsphysiologie* 10:504-15; 1939.
77. KARP, OTIS. "Action Potential Analysis of Golf Drive." Master's thesis. Urbana: University of Illinois, 1955.
78. KISTLER, J. W. "Study of Distribution of Force Exerted upon the Blocks in Starting from Various Positions." *Research Quarterly Supplement* 5:27-34; 1934.
79. KLOPSTEG, PAUL E.; WILSON, P. D.; and OTHERS. *Human Limbs and Their Substitutes.* New York: McGraw-Hill Book Co., 1954.
80. KRAKOWER, HYMAN. "Skeletal Symmetry and High Jumping." *Research Quarterly* 12:218-27; May 1941.

81. Leksell, L. "The Action Potential Excitatory Effects of the Small Ventral Root Fibers to Skeletal Muscles." *Acta Physiologica Scandinavica Supplement* 10; 1945.
82. Lloyd, D. P. C. "Facilitation and Inhibition of Spinal Motoneurones." *Journal of Neurophysiology* 9:421-38; 1946.
83. Luhr, Overton. *Physics Tells Why.* Lancaster, Pa.: Jaques Cattell Press, 1944.
84. McCloy, C. H. "The Organization and Teaching of Apparatus Work and Tumbling." *Journal of Physical Education* 34:60-62; 1937.
85. Magoun, H. W. "Caudal and Cephalic Influences of the Brain Stem Reticular Formation." *Physiological Review* 30:459-74; 1940.
86. Marey, E. J. *Animal Mechanism.* New York: D. Appleton-Century Co., 1887.
87. Mathews, B. A. C. "The Response of a Muscle Spindle During Active Contraction of a Muscle." *Journal of Physiology* 72:153-74; 1931.
88. Minkowski, M. "Sur les mouvements, les réflexes, et les réactions musculaires du foetus humain de 2 à 5 mois et leurs relations avec le système nerveux." *Revue Neurologique* 37:1105-18, 1235-50; 1921.
89. Morton, Dudley J., and Fuller, D. D. *Human Form and Locomotion.* Baltimore: Williams and Wilkins Co., 1952.
90. Nakamura, H. "Experimental Study of Reaction Time of the Start in Running a Race." *Research Quarterly Supplement* 5:33-45; 1934.
91. Penfield, Wilder. "Mechanisms of Voluntary Movements." *Brain* 77:1-17; 1954.
92. Ralston, H. J. "Mechanics of Voluntary Muscle." *American Journal of Physical Medicine* 32:166-84; 1953.
93. Rehling, C. H. "Analysis of Techniques of Golf Drive." *Research Quarterly* 26:80-81; March 1955.
94. Rehman, I. "A Study of the Kinematics and Dynamics of the Human Gait and Its Application in Poliomyelitis." *Archives of Physical Medicine* 28:749-56; 1947.
95. Rehman, I.; Patek, P. R.; and Gregson, Margaret. "Some of the Forces Exerted in the Normal Human Gait." *Archives of Physical Medicine* 29:698-702; 1948.
96. Richet, M. P. "Note sur la contraction du muscle quadriceps dans l'acte de donner un coup de pied." *Comptes Rendus de la Société de Biologie* 2:204-205; 1895.
97. Rork, R., and Hellebrandt, F. A. "The Floating Ability of Women." *Research Quarterly* 8:19-27; December 1937.
98. Schwartz, R. P.; Heath, A. L.; and Wright, J. W. "Kinetics of the Human Gait." *Journal of Bone and Joint Surgery* 16:343; 1934.
99. Seyffarth, H. "The Behavior of Motor Units in Voluntary Contraction." *Physiological Foundations of Neurology and Psychiatry.* (Edited by Gellhorn.) Minneapolis: University of Minnesota Press, 1953.
100. Sherrington, Charles S. The Integrative Action of the Nervous System." New Haven: Yale University Press, 1923.
101. Shevlin, P. O., and McCloy, C. H. "Electromyographic Study of Selected Muscles Involved in Movements of Upper Arm at Scapulohumeral Joint." *Research Quarterly* 27:409-17; December 1956.
102. Slater-Hammel, A. T. "Possible Neuromuscular Mechanism as a Limiting Factor for Rate of Leg Movement in Sprinting." *Research Quarterly* 19:745-56; December 1941.
103. Slater-Hammel, A. T. "An Action Current Study of Contraction-Movement Relationships in Golf Strokes." *Research Quarterly* 19:164-77; October 1948.
104. Slater-Hammel, A. T. "Action Current Study of Contraction-Movement Relationships in Tennis Stroke." *Research Quarterly* 20:424-31; December 1949.
105. Slocum, D. B. *An Atlas of Amputations.* St. Louis: The C. V. Mosby Co.: 1949.
106. Sperry, R. W. "The Functional Results of Crossing Sensory Nerves in the Rat" *Journal of Comparative Neurology* 78:59-90; 1943.

107. SPERRY, R. W. "Action Current Study in Movement Coordination." *Journal of General Psychology* 20:295-313; 1939.
108. STEINDLER, A. *Mechanics of Normal and Pathological Locomotion in Man* Springfield, Ill.: Charles C. Thomas, 1935. 424 p.
109. STETSON, R. H., and BOUMAN, H. D. "The Coordination of Simple Skilled Movements." *Archives néerlandaises de physiologie de l'homme et des animaux*, 20:179-254; 1935.
110. STETSON, R. H., and BOUMAN, H. D. "The Action Current as Measure of Muscular Contraction." *Science* 77:219-21; 1933.
111. SZENT-GYORGYI, A. *Chemistry of Muscular Contraction.* New York: Academic Press, 1953.
112. TUTTLE, W. W.; MOREHOUSE, L. E.; and ARMBRUSTER, DAVID. "Two Studies in Swimming Starts." *Research Quarterly* 10:89-98; March 1939.
113. TUTTLE, W. W.; MOREHOUSE, L. E.; and ARMBRUSTER, DAVID. "A Further Study of the Optimum Time for Holding a Swimmer on His Mark." *Research Quarterly* 11:53-56; March 1940.
114. WALKER, G. A., and HAYDEN, T. C. "The Optimum Time for Holding a Sprinter Between the Set and the Stimulus (gun shot)." *Research Quarterly* 4:124-30; 1933.
115. WALTERS, C. ETTA, and PARTRIDGE, MARIAM J. "An Electromyographic Study of the Differential Action of the Abdominal Muscles." *American Journal of Physical Medicine* 36:259-68; October 1957.
116. WATSON, KATHERINE C. *A Study of the Relation of Certain Measurements of College Women to Throwing Ability.* New York: New York University, 1935.
117. WEBER, E., and WEBER, W. *Mechanik der Menschlichen.* Gehwerkzeuge: Gottingen, 1836. Quoted by Hans Bollin. *Mind and Body* 1, 1894.
118. WEISS, P. A. "Self Differentiation of the Basic Patterns of Coordination." *Comparative Psychology Monograph* 17:1-96; 1941.
119. WESTERLUND, J. H., and TUTTLE, W. W. "Relationship Between Running Events in Track and Reaction Time." *Research Quarterly* 2:95-100; October 1931.
120. WHITE, R. A. "Effect of Hip Elevation on the Starting Time of the Sprint." *Research Quarterly Supplement* 6:128-33; 1935.
121. WILD, MONICA. *The Behavior Pattern of Throwing and Some Observations Concerning Its Course of Development in Children.* Doctor's thesis. Madison: University of Wisconsin. 1937.
122. WINDLE, WILLIAM F. "Reflexes of Mammalian Embryos and Fetuses." *Genetic Neurology.* (Edited by Paul Weiss.) Chicago: University of Chicago Press, 1950. p. 214-22.
123. ZIMMERMAN, HELEN M. "Charecteristic Likenesses and Differences Between Skilled and Non-Skilled Performance of Standing Broad Jump." *Research Quarterly* 27:352-62; October 1956.

Methods of Instruction

M. GLADYS SCOTT

Why should the researcher be concerned about teaching method? Has not each teacher received the traditional instruction on how to present activity and develop skill in his respective specialties? Is not the good teacher alert to success and lack of it and thereby through experience is he not developing an optimum method for himself? Does not the learning process remain unchanged even though this is a changing world?

All the above points may be true to a certain extent. However, teachers are faced with ever increasing numbers of students and, in general, larger classes. The work of the educator is being scrutinized as never before, and any area of education which does not have a clear-cut goal and an efficient method of working toward it is going to find itself deleted or under extra pressure to produce results.

The teaching of physical education today varies from excellent to poor. Some teachers are not even aware of what their results are and how much improvement could be made. They are so conditioned by habit to certain practices that they just do not try new approaches.

The best results can be achieved only in terms of continued critical evaluation of experience, use of all factual and other data, and, above all, objective research in pursuit of the technological application of recognized principles on modification of human behavior. Research is an important ingredient—the leavening—for this process of improved teaching method. Therefore, it is important that every teacher work directly in this process of study and experimentation on learning.

Unfortunately, too, there is an insufficient amount of objective evidence for us to rely upon as a substitute for this self-help. Kretchmar (10:245), writing in 1949, said, "Research has barely begun on the heart of learning and teaching problems." The work of another decade has not changed that situation significantly. *Method* is here construed to mean not only the process of presenting concepts and guiding the learner to some reaction to the situation, but the whole process of organizing the learning environment, of determining the curricular goals, and of planning the sequence of experiences which we sometimes refer to as progression.

Let us consider some of the problems which might be studied.

INDIVIDUAL DIFFERENCES

Teachers learn about the existence of individual differences but what do they do about them? An excellent review of assessing individual differences can be found in Frandsen (6: Chapters 12 and 13). The schools have accepted the principle of the "opportunity room," remedial reading, remedial speaking, special schools for the physically handicapped, special teams, and opportunities for the highly skilled. Why not special classes for the "motor retardees"? They too can be helped. The efforts to date have been primarily at the college level, as for example Lafuze (11) and Broer (1). Persons with low motor skills should be recognized early and given special help. Instruction should be on a preventive rather than remedial basis.

Lafuze (11) has demonstrated that separate classes for the less skilled taught in the same way as for the more skilled are not the best method. It is necessary to adapt teaching techniques to the level of ability. Further work should look toward optimum methods for these slow learners.

The research steps here can include a creative analysis of needs and development of modified learning experiences. Then a controlled experiment such as is outlined in Chapter 10 will determine the merits of the teacher's hunches, dreams, and new ideas.

Another approach to individual differences is by way of measurement and objective determination of various abilities. Tests may be selected from the literature or developed to meet a specific need according to the method outlined in Chapter 8. Again, in development of new measuring devices, the teacher's creativity is challenged.

After capacities of the individual students have been measured, various steps may be used for effective follow-up efforts on learning. Achievement scales may be constructed, or profiles prepared for each student, or graphs drawn for the performance of the class. (See Chapter 7.)

The possible processes of interpreting the test results to the students may be appropriate variables for another controlled experiment. The reaction or attitudes of the learner may provide an opportunity to use the attitude scales (Chapter 5) or sociometric interaction as a part of the outcomes of one or more of the above experiments.

Additional studies would be helpful comparing the learning rate of groups which differ in learning capacity. Equally valuable would be a comparison of effectiveness of a given method, of optimum length of lessons, of frequency of practice, of socio-metric status, of associated capacities and interests of these diverse groups.

METHODS OF TEACHING-LEARNING

The classroom, which should be the site of constructive experimentation, is a realistic laboratory where the variables of method may be applied one at a time and with appropriate subjects. These variables may be those based on psychological precepts, or variations in practice and experience, or means of acquiring concepts and motor co-ordinations as expressions of those concepts. Classroom experimentation is a necessary supplement to experiments on animal or human learning in the laboratory on selective tasks as individual learners.

The psychological guides or laws of learning lead to establishment of different methods of teaching. The classic debate over the merits of the part method vs. the whole method is well known. It leads on the one hand to a premise that drill on carefully selected details, to be put together after perfection of each, is the route to learning. It is in contrast to the premise that a skill is a unfied whole to be comprehended mentally, tried until the complex pattern emerges as a co-ordinated, functional performance. This illustrates the importance of a working hypothesis, a philosophical foundation on which the creativity of the design springs up.

Studies may be found on either the part or the whole method, or various adaptations of one or the other. Part learning tends to develop ability on each part, but in the case of a complex skill there is usually difficulty in trying to put the parts together. Knapp and Dixon (9) showed the whole method superior in learning to juggle. Lambert (12) found interference from learning each hand separately on a two-handed manipulatory skill. The logical conclusions would seem to be that the appropriate method will vary with the complexity of the skill to be learned and with the mechanics of the skill to be learned.

The working hypothesis would seem to be based on answers to questions such as the following. These should be carefully considered by any investigator in the realm of part-whole methods.

1. What is an identifiable whole?
2. Is that whole mechanically possible in separate parts?

3. What alterations are made by dividing it into parts?
4. What similar skills does the subject know which may serve as a substitute concept and neuromuscular pattern?
5. Are all ability levels going to respond in the same way to the units or parts or whole presented?
6. What units will the subjects adopt as working parts, if it is presented as a whole?
7. Can the parts be related better mechanically in a series of definite parts or by progressively adding parts as the previous one is learned?
8. Will the optimum method be the same regardless of whether the time invested by each subject in the learning process is small or great?
9. Will the optimum method be the same for skills of simple and complex nature?

Answers to the above questions may not be possible from the literature in its present stage or on *a priori* basis. However, careful weighing of these questions will certainly lead to better planning of experiments (Chapter 10) in this area and to a more rapid approach to solution of some of our queries on part-whole learning.

The procedures just outlined will probably lead to evidence that different activities can be learned best by different approaches. This is largely a matter of complexity of the skill. For example, shuffleboard and bowling might be considered similar in form, both requiring accuracy. However, the additional requirements of accuracy and timing in bowling make it a much more complicated motor task. Or, an elementary backstroke and a crawl stroke differ in complexity, so that methods for these and other strokes may differ.

Experiments tend to be conducted on relatively few subjects and in small groups. It seems very probable that some differences may exist in desirable methods and learning in large classes. Adequate study has not been done, using controlled experiments in classes of different sizes, on ways of improving instruction in large classes. For example, it is known from experience that some assistance can be obtained by using student aides; developing student habits of self-analysis and help; using still pictures, motion pictures, films and filmstrips; establishing objective goals toward which to work; and other forms of extending the teacher's services. The most valuable techniques could be identified by controlled experiments. (See Chapter 10.) Administrators under pressure of numbers of students really become enthusiastic when evidence can be presented for conclusions such as this, "Class achievement is very similar to that of small groups." (17:98)

Perhaps one way to extend the opportunities for learning afforded by the gymnasium is to develop techniques of mental practice on skills being learned. There seems to be some evidence that it is helpful (18). Questions which cannot be answered are the stage or stages in the learning curve where mental practice is beneficial: how much is required for learning to occur, at what point the learner is wasting time, what assistance can be given the learner to develop effective habits of mental practice, and what relationship, if any, exists between kinesthetic acuity and ability to profit from mental practice. Profitable studies in this area could be numerous. Some could be exploratory and accomplished through measurement, with co-relational and careful analytical description of data. Most would need to be conducted by means of the controlled experiment in relation to classroom (gymnasium) learning. (See Chapter 10.)

LEARNING AIDS

Best known among the teaching aids are the audio-visual materials. Sound films are well validated as teaching devices. They may be used for explanatory demonstrations (8), for contrasting right and wrong techniques (7), for testing purposes (7), or for incentive and motivation, as well as being bearers of information (13). Less well known are the loopfilms, filmstrips, still pictures, charts, graphs, models, illustrative card file, self-written descriptions by class members, and other ingenious ways of learning through multiple sensations. Ragsdale (15) expresses this point of view in writing concerning the processes of motor learning. He says that not one, but all, the sensory channels for learning must be used.

Damron (4) used a tachistoscopic training device and found little difference in value between the two- and three-dimensional materials.

Ruffa's study (16) on use of films led him to conclude that their use results in more independent learning and more self-confidence in what the learner was doing. He says that the use of films for explanation of motor acts implies a certain amount of imitation, which serves to guide the learner. However, all learners may not profit equally from trying to imitate, and Ruffa believes this may depend upon ability to make symbolic representation. Can this symbolic representation also include kinesthetic interpretation of what is seen and empathetic experience during the observation? This possibility could well be the basis of a study on relationship of kinesthetic perception and use of observation and imitation in teaching. Deese (5) gives rules for the use of imitation. He says one must make very clear what is to be imitated; and to try to imitate short, simple acts only. Furthermore, the act to be imitated must be presented in such a way that the learner does not need to change any part of it in order to perform (5: 232). These might well be guides in setting up studies relative to imitation.

Basically, it is a question of *how* films should be used, not whether they should be used. The same could probably be said of all other forms of the learning aids. There are appropriate problems here for investigation of optimum usage.

ATTITUDES

In considering methods of instruction, two aspects of the problem of attitude seem pertinent. On the one hand, attitude determines the receptivity of students to proposed learning opportunities. On the other hand, attitudes are modified, created, and nurtured by the learning experience. Classroom learning is more or less guided by the activities and efforts of an instructor, so these attitudes may encompass the teacher and his processes of teaching as well as the activity being learned, physical education in general, and even at times the whole process of education.

The first requirement of learning is to have the learner in a receptive, happily anticipatory mood. His mind is set for action, for work, for accomplishment. The challenge for the teacher then is to provide meaningful practice, hard work, recognition of achievement, and the fun of working together and mutually respecting each other's accomplishments. Therefore, research in this interrelated attitude-learning situation is based on hypotheses such as the following:

1. Attitudes are measurable or identifiable.
2. Attitudes are modifiable.

3. Attitudes are not static; they improve or deteriorate.
4. Attitudes are an inevitable accompaniment of any situation. Humans do not live or act in a detached fashion within their environment.

Studies on measurement of attitude have tended toward attitude scales or sociometric analysis (Chapter 5) or projective tests (Chapter 5). The projective tests recognize subconscious controls over attitudes and reactions and attempt to obtain a "candid shot" of what is really making the person react as he does. The sociometric analysis recognizes the group dynamics which are so important in a class situation, particularly in physical education where the students attempt to work and play together so continuously.

Research is greatly needed on refinements of attitude scales. Like all tests, the need is for short forms made up of critica or valid items, which the student will accept as a procedure and on which he will co-operate. The work on projective measures and sociometrics needs to include both physical education oriented cues and also devices for improving the unrestrained, revelatory responses. With improved techniques with which to work, then the controlled experiment can deal with the variables of method, work intensity, knowledge of accomplishment, reaction to drill, coeducational classes, levels of perfection in goals established, and many other problems of method and curriculum.

For example, the discrepancy between teacher and student goals is often the source of frustration for both. What level of skill do we want, practical or expert? Ragsdale (15:71) says that learners often prefer to stop trying to improve at a *practical* rather than an *expert* level. The practical level is the one which represents enjoyment, satisfaction in results in terms of effort spent, and a real leisure-time course of enjoyment. The expert level requires serious, consistent work; for the majority it becomes work, not play, and fails to yield satisfaction. Teaching must approach complex skills without an excessive, fun-destroying array of details and criticism. Therein lies the challenge for a succession of studies at different age levels, of different activities, and for both sexes.

Student goals are usually concerned with use of skills, not just practice, and as they say, "playing and having a good time." Miller and Ley (14:5) very aptly state a plausible bridge between the student and teacher's goals:

> "Students can begin to play almost as soon as they begin to learn the skills of a game. It is not good teaching procedure to attempt to perfect a student's skill before putting it to use, nor is the perfection of one skill necessary before other skills are presented. The good teacher must recognize individual limitation and through the use of sane progression and logical arrangement in the presentation of both skills and playing techniques, enable every student to enjoy playing the game while she is developing the skill."

Again this serves as a working hypothesis on which to build a series of experiments and a challenge to the creativity of many instructors.

Another aspect of attitude has to do with safety during the learning process and during later use of acquired skills. Experience has shown that the learner is safe as long as he is working within the range of his ability and working with confidence. What is said to a performer makes him hesitant and uncertain, or overly reckless, or calm and confident. The teacher can make or break a gymnast, a swimmer, a dancer, or a football player. This is a phase of teaching where the teacher can accept the challenge to reduce accidents to the minimum and increase learning to the maximum. The evaluation of results will come in charting the progress of these two divergent records and in comparing the skill and attitudes under the more favorable learning climate.

Brown (3) says that fear is learned . . . and notes its internal manifestations. All teachers have seen its effect on the learning process. Swimming is probably the most common example and much has been learned in this area which will apply in other activities. For example, how is a safety pole used in swimming? Is it a crutch which deprives the swimmer of the experience of really taking care of himself, or is it an aid which he knows will be used in case of real necessity? Is it used in such a way that it means to the swimmer, "I don't think you are going to succeed so I am ready with the pole over your shoulder," or "I am sure you can make it, I am watching and will help." These are subtle differences but they affect the acquisition of skill, development of attitudes, and development of self-sufficiency in the water. The latter is built on the premise of carefully planned progressions, encouragement of self-confidence, recognition of ability, and quiet unobtrusive aid. This premise invites informal experimentation by the teacher, and controlled experimentation by the teacher interested in research. The same situation exists in tumbling, apparatus and trampoline, dance of all types, and even for some children in the matter of learning to catch a ball or to kick or bat it.

And so whether teachers are aware of it or not, the gymnasium, playing field, and dance studio become the laboratory in which the elements of pupil interests, the teacher's planned experiences, group interaction, other school experiences, and cultural pressures unite to synthesize into learning, understanding, and attitudes—ultimately to develop our citizens of tomorrow. No competent chemist would turn his back, use the chance method, or even use the method of common practice. He is constantly following methods prescribed from previous studies and constantly experimenting for better techniques. The teacher, as a technologist, should work in his own special laboratory with the same scientific curiosity, discipline, and technique. The reward is a healthier, happier, more accomplished student. But what is imperative is that as professional persons, teachers have no other course.

BIBLIOGRAPHY ON METHODS OF INSTRUCTION

1. BROER, MARION R. "Evaluation of a Basic Skills Curriculum for Women Students of Low Motor Ability at the University of Washington." *Research Quarterly* 26:15; March 1955.
2. BROWN, CLARENCE W., and GHISELLI, EDWIN E. *Scientific Method in Psychology.* New York: McGraw-Hill Book Co., 1955. 368 p.
3. BROWN, JUDSON. *Current Theories of Research in Motivation. A Symposium on Current Theory and Research in Motivation.* Lincoln: University of Nebraska Press, 1953.
4. DAMRON, C. FRAZIER. "Two- and Three-Dimensional Slide Images Used with Tachistoscopic Training Techniques in Instructing High School Football Players in Defenses." *Research Quarterly* 26:36; March 1955.
5. DEESE, JAMES. *The Psychology of Learning.* New York: McGraw-Hill Book Co., 1952.
6. FRANDSEN, ARDEN N. *How Children Learn.* New York: McGraw-Hill Book Co., 1957. 546 p.
7. HOMEWOOD, JEAN M. "A Film and Manual for Use in Developing Ability to Teach Sports Skills." Doctor's thesis. Iowa City: State University of Iowa, 1955.
8. KERN, REBECCA W. "The Value of Visual Aids in Teaching Elementary Diving to College Women." Master's thesis. Iowa City: State University of Iowa, 1951.
9. KNAPP, CLYDE G., and DIXON, W. ROBERT. "Learning to Juggle: II. A Study of Whole and Part Methods." *Research Quarterly* 23:398; December 1952.
10. KRETCHMAR, ROBERT J.; SHERMAN, HOYT; and MOONEY, ROSS. "A Survey of Research in the Teaching of Sports." *Research Quarterly* 20:238; October 1949.

11. LAFUZE, MARION. "Learning of Fundamental Skills by College Freshmen Women of Low Motor Ability." *Research Quarterly* 22:149; May 1951.
12. LAMBERT, PHILIP. "Practice Effect of Non-Dominant vs. Dominant Musculature in Acquiring Two-Handed Skill." *Research Quarterly* 22:50; March 1951.
13. LUTTGENS, KATHRYN. "The Construction of a Film for Use in Instruction in Posture and Body Mechanics." Doctor's thesis. Iowa City: State University of Iowa, 1957.
14. MILLER, DONNA MAE, and LEY, KATHERINE L. *Individual and Team Sports for Women.* Englewood Cliffs, N.J.: Prentice-Hall, 1955.
15. RAGSDALE, C. E. "How Children Learn the Motor Type of Activities." Forty-Ninth Yearbook, Part I, *Learning and Instruction.* National Society for Study of Education. Chicago: University of Chicago Press, 1950. p. 69
16. RUFFA, EDWARD. "How Motion Pictures Help the Coach." *Athletic Journal* 17:20; May 1937.
17. SCOTT, M. GLADYS. "Learning Rate of Beginning Swimmers." *Research Quarterly* 25:91; March 1954.
18. TWINING, WILBUR E. "Mental Practice and Physical Practice in Learning a Motor Skill." *Research Quarterly* 20:432; December 1949.

Applied Physiology

MARJORIE U. WILSON
JOSEF BROZEK, CONSULTANT

Applied physiology, as used here, refers to experimentation in the field of physiology which results in observations directly related to the human organism. Special emphasis is placed on the reactions of man to his environmental stresses. Physical education is concerned primarily with the physiology of exercise and work as modified by climate, nutritional factors, and changes with age and sex differences. Hence, there has been a rapid growth in the development of research laboratories in physical education departments in colleges and universities. These may range in pretentiousness from a single room with a few pieces of equipment to well-equipped laboratories staffed with trained technical and professional researchers.

For the physical educator, the scope of research in these areas is limited by the frequent demand in modern physiological research for highly trained personnel and elaborate laboratory equipment. However, no student should deter because he does not have large sums of money to spend for equipment. Even the gymnasium or the playing field can become a creditable natural laboratory.

Physiology of exercise is a broad area that is concerned with the performance of muscular work under various conditions. It also is concerned with the underlying physiological mechanisms that make such performance possible. Individual differences in functional ability and in efficiency, as well as group comparisons, are of interest. Studies relative to endurance, fatigue, and measures of energy output hold special significance for the student of physical education. The scope of this section will not permit complete coverage of the many topics that logically fall within the limits of laboratory physiology. The purpose here is to present a guide to some of the typical methods of physiological research that can be conducted by the physical educator.

No specific methodology can be discussed at length in this chapter. Instead, selected references will be cited to illustrate the problems in question. Recency of reference is therefore of less importance than appropriateness. Experimental design, per se, is discussed in Chapter 10 and statistical procedures are presented in Chapter 7.

A discussion of research in applied physiology could be approached in several ways. Here, three topics will be discussed: the methods used to obtain measurements, standardization of procedures and conditions, and experimental problems.

METHODS OF MEASUREMENT

For some purposes it may be desirable to study the athlete in his natural state of activity; for example, to examine the response of his cardiovascular system and the energy expenditure while swimming or playing basketball. This can be done, under certain circumstances, by, using special techniques, including telemetering of the raw data. However, rigorous experimentation during performance is not usually feasible. In the laboratory the experimenter may simulate a particular athletic performance situation or, more likely, he may select a standardized type of work, such as treadmill-walking or riding a bicycle. Under the conditions of standardized performance, numerous physiologic measures can be made with precision.

Mechanical Work. The amount of physical work that can be performed by the human machine under diverse conditions is of great interest to many individuals—nutritionists, industrial employers, physicians, coaches, and physical educators. This simplest form of work is defined by the physicist as the product of the force applied and the distance through which this force acts. The lifting of the human body against gravity through a certain height or the pushing of an object horizontally for a specified distance is called *mechanical, external,* or *visible* work. The customary units of measurement for external work are foot-pounds or kilogram-meters.

The use of the step test as a prescribed unit of work is a direct application of the principle of mechanical work. There are obvious discrepancies in such a calculation. No work is recorded for lowering the body to starting position. Swinging of arms and legs, starting and stopping, stabilization of parts of the body are all excluded from the measure of total work done. Consequently, it is impractical to use the step test when a finite amount of work is to be assessed. The step test can be justified philosophically on the basis that the subject is lifting his own weight and handling his own body, a feat which he must perform regularly in any physical activity. Tuttle (34) used the step test to develop his pulse-ratio test. The Harvard Step Test (1) used stool stepping to define work done in a measure of physical condition. Weight lifting and stair climbing also have been used as measures of external work. Here, again, the weight lifted through a specific height furnished the estimate of work done, but no attempt was made to measure output on the return movement.

Since it is difficult to measure work output of an isolated muscle or even groups of muscles in man, investigators have turned to the other extreme—an attempt to measure the mechanical output of the majority of the muscles of the total body. One device for this kind of measurement is the bicycle ergometer in which the subject usually pedals against resistance—a weight, friction resistance, magnetic brakes, or an electric generator. Careful calibration of bicycle ergometers makes a rather precise measure of external work feasible. There are numerous criticisms of such an ergometer: (a) The type of work done is artificial; (b) The position assumed may interfere with ventilation and circulation; (c) The greater share of work is done by the legs; (d) The load (resistance) is assigned arbitrarily without consideration for individual differences in strength and endurance; (e) Preliminary training is necessary to insure reliable results. In spite of criticism, the bicycle ergometer is a popular means of controlling the amount of work done in physiological experimentation. Tuttle (29, 36), Hellebrandt (11, 19), Karpovich (17, 18), and many others have employed this type of work in their investigations.

The treadmill, another device to measure total external work, is preferred by many experimenters because the work is done in a more natural position and because the activity of walking does not have to be learned. Other desirable factors usually mentioned are the automatic pacing for the subject and the natural weight load. The cost and size of the equipment limit its use to the large, well-equipped laboratory.

These relatively refined measures of external work may be used to control a quantity of work imposed upon experimental subjects or to determine the subjects' capacity for work as conditions are varied. Other less accurate measures of external work have been used because they demand little or no equipment. At times the selection of calisthenics, running, and swimming may be preferred because the accurate measure of work done appears less important than naturalness of the activity. The experimenter should be cautioned not to sacrifice precision of measurement for convenience.

Physiological Work. The total amount of energy consumed in performance of a task can be measured directly by measuring the amount of heat produced, or indirectly by measuring the amount of oxygen consumed. These measures are usually called *physiological work* and the method is classed as *calorimetry*. Direct calorimetry measures the total heat loss of an individual over a given period of time by placing him in a closed respiration chamber. The heat from the subject is absorbed by the water and air surrounding the chamber; the differential is electrically recorded. Equipment and staff required to operate this method, as well as the relative confinement of the subject, make it impractical for use in investigation of most motor activities.

Indirect calorimetry, a method of measuring oxygen consumption, is a customary way of obtaining total calories produced. Thence, the approximate caloric production can be calculated from the rate of oxygen consumption. The experimenter can select either of two acceptable approaches to this method—the closed circuit or the open circuit technique.

With the closed circuit, the subject breathes pure O_2 from a specially designed spirometer and expires into a soda lime container where the CO_2 is absorbed. Movements of the spirometer cylinder are recorded on a revolving drum. The rate of respiration, tidal volume, and amount of O_2 consumed can be determined directly from the kymographic record. In the open circuit technique, the subject breathes atmospheric air and exhales into a reservoir for gases, such as the Tissot spirometer or the Douglas bag. The Douglas bag is the preferred receptacle for many studies involving movement in space; it is portable and can even be carried while the subject is exercising (37). The expired air, for a specified period of time, is measured and the gas analyzed for O_2 and CO_2 content. The latter method is time-consuming, but the results are considered more accurate because gas analysis apparatus is highly sensitive.

Two relatively new respirometers, which are unique because of their portable feature, may open the way to increased physiological research by physical educators—the Kofranyi-Michaelis calorimeter (21, 26) produced in Germany and the Integrating Motor Pneumotochograph (7) produced in France. Both respirometers measure the exhaled volume directly and store an aliquot of each tidal volume in a bladder for later analysis. Either machine can be worn on the back as a pack, and contributes little to the work load as their total unit weighs about 8 pounds or less. Either machine should be applicable to all forms of activity except contact sports.

To calculate energy costs, three measures of O_2 consumption must be made: (a) O_2 consumption during rest, (b) O_2 consumption during a specified work period, and (c) O_2 consumption during recovery after exercises. The addition of the work O_2 and recovery O_2 minus the resting O_2, when converted to equal time units, results in a net O_2 cost per unit time. The intensity of energy expenditure can be expressed in the ratio of work rate to resting metabolic rate. Since, for every liter of O_2 consumed, approximately 5 calories of energy are liberated, O_2 cost can be expressed in calories. For a more accurate caloric equivalent, the RQ (the ratio of CO_2 produced to O_2 consumed) would need to be determined for each subject.

In some athletic events, like team sports and track events, it is impractical to use a Douglas bag to determine energy costs. Weiss and Karpovich (38) have used a circuitous approach based on the relationship between oxygen used and pulse rate per minute. A subject's pulse rate and oxygen consumption is recorded for work that can be measured (treadmill or bicycle ergometer) at various intensities. These measures are graphed for a given subject. Since there is a linear relationship between these two measures, if the pulse rate per minute is known for this subject his oxygen consumption thereafter can be calculated.

Efficiency. The efficiency of the human machine is expressed as a ratio of external work done to amount of energy used in performance of that work, which is usually converted to a percentage.

$$\text{Gross efficiency} = \frac{\text{mechanical work}}{\text{total physiological work}} \times 100$$

Since a certain amount of energy is expended constantly just to maintain physiologic processes necessary to life, gross efficiency does not give a true picture with respect to any given work task. The energy used in maintenance of body processes can be subtracted from the total physiological work to produce a measure of net energy used.

$$\text{Net efficiency in percent} = \frac{\text{energy equivalent of mechanical work (calories)}}{\text{energy equivalent of total physiological work minus resting energy expenditure}} \times 100$$

Ventilation. The need for measures of respiratory flow in human subjects is apparent because work metabolism is dependent upon oxygen consumed. The rate of respiration and tidal volume, as previously mentioned, can be determined from the kymographic record in closed circuit calorimetry. The pneumograph provides another method of recording respiratory movements. It consists of an elastic tube or inflated elastic bag, strapped firmly about the chest, which is connected by means of rubber tubing to a tambour or drumlike membrane. Change in chest size alters the pressure in the tube around the chest which in turn stretches or recoils the membrane of the tambour. A lever on top of the tambour records these excursions on a revolving drum. Calibration is essential if volume measures are to be considered. It is questionable whether absolute thoracic volume changes can be shown by this or any other similar device that measures external respiratory movements. Any extraneous movements would also appear on the record. The body plethysmograph, another method of measuring rate and amplitude of ventilation, is discussed by Greulich (8). Many more technical procedures of measuring respiratory flow, volume, composition, and capacity are discussed in detail in *Methods in Medical Research* (3: Vol. 2, Section 2). Vital capacity, a measure of total respiratory capacity, is generally measured by asking the subject to expire as fully as possible into a spirometer through a rubber tube. This measure is so unreliable that it is customary to select the highest score out of ten trials as a measure of vital capacity. References to original papers on lung volume and the efficiency of pulmonary ventilation are given by Greulich (8) and by Peters and Van Slyke (28).

Blood and Circulation. For a number of reasons, the examination of blood samples taken from human subjects is still considered a clinical problem and, under most circumstances, is beyond the scope of the physical educator. At present two approaches based on the dilution principle are available for the estimation of blood volume. One approach involves the determination of the plasma volume by the intravenous injection of a known amount of a harmless dye or radio-iodinated plasma protein. The second approach estimates the red cell volume and can be done in two ways. The first requires the intravenous injection of red blood cells labeled with radioactive elements or red blood cells of a group compatible with, but different from, that of the subject. The second requires that the subject breathe a known amount of carbon monoxide which is taken up by the hemoglobin of the red cells to form carboxyhemaglobin. When a simultaneous determination of the hemotocrit is combined with the estimation of the plasma or red cell volumes, the blood volume may be calculated. There is a third approach for blood volume estimation, termed electrical impedance, which is not dependent upon the dilution principle. This method depends upon the measurement of the alterations in the conductivity of blood resulting from the intravenous injection of hyper- or hypo-tonic solutions. So far, this method has not been applied to human subjects.

Further discussion of methods that can be used to determine approximations of blood and plasma volume by the carbon monoxide technique or the dye technique are found in Greulich (8), in Peters and Van Slyke (28), and in Keys and others (20: Vol. 2, Appendix 1). References to original papers are included in these sources. In the above three references and *Methods in Medical Research* (3: Vol. 2, Section 2), numerous methods in areas of blood composition, O_2 saturation, and acid-base balance are developed in detail.

Blood pressure determinations, as well as pulse rate and pulse recovery, have been accepted as standard procedures in the measure of physical condition. The study of post-exercise and post-experimental blood pressure and pulse determinations by Happ (10) and Moutis (25) serve as examples. Venous pressure provides an indication of the efficiency with which blood is returned to the heart. A typical indirect method of measurement is described by Eyester (5). A more satisfactory method appears in *Human Starvation* (20: Appendix 1). *Methods in Medical Research* (3: Vol. 2, Section 2) describes clinical methods of measuring blood flow.

On occasion, experimenters cannot find adequate apparatus to investigate the problem at hand. Then adaptation or modification of a given test or piece of equipment may be necessary before research can be undertaken. At times this process is sufficiently extensive to provide a study in itself, such as the modification and calibration of the bicycle ergometer by Tuttle and Wendler (36) and the development by Henry (13) of the electrical cardiometer, a device for recording heart rate, number of step-ups, and time simultaneously. The development of new tools of measurement is important not only to ensure more precise physiological measurement but also to provide a means of study in areas that previously eluded measurement. Tuttle and others (35) designed and constructed new types of dynamometers, based on the strain gauge technique, which provide a more accurate measure of strength and also a measure of strength endurance. Simonson and associates (31) developed an apparatus called a ballistic elastometer, a dynamic method for estimating elastic properties of skeletal muscle. Adequate experimentation with new apparatus to assure standardization and simplification of operational procedure, in order that others may successfully utilize the equipment, is an integral part of such a study. The *Journal of Applied Physiology* has a section called "Special Communications" which is devoted to new methods of research.

The development of new physiological performance tests and/or the validation of such tests as conditions are altered in the administration are important to methodology. Brouha (1) developed a simple step test and evolved formulas by which pulse rates after exercise can be predicted for varying levels of training. Since it is natural for experimenters to vary conditions of a given test to fit their immediate needs, Elbel and Green (4) studied the effect of changing bench heights and Miller and Elbel (24) investigated the effects of change in tempo in step tests. All of these represent studies in the areas of applied physiology.

STANDARDIZATION OF CONDITIONS

Research in the area of applied physiology is typically experimental in nature. The fundamental purpose of such research is to determine the effect on human subjects of selected variables in areas of training, nutrition, work output, fatigue, etc. The need for precision of measurement is self-evident. The degree of precision possible is governed to a large extent by the success with which factors—other than the deliberate variations included within the predetermined experimental design—can be and are controlled or equalized. Complete and absolute precision in human measurement is impossible because:

1. The sampling is usually necessarily small when numerous determinations must be taken.
2. Human effort is influenced by numerous factors that cannot be controlled or factors that are too vague to be identified.
3. No measure can be more precise than the inherent accuracy of the measuring instrument.
4. Human fallacy of the experimenter may be a source of error.

It is therefore imperative that all extraneous factors be controlled or, if this is impossible, that all results of the study be qualified accordingly (33).

Basic measures must be carefully taken before experimental variables are introduced. Enough of these resting state or "pre-experimental" state measures need to be taken to assure validity of measurement. The experimenter must be certain that no circumstance prior to experimentation, such as physical exertion, smoking, illness, or drugs, has in any way influenced the physiologic measures to be applied to his subjects. For example, a "resting" pulse rate or O_2 consumption rate taken at the beginning of an experiment may be influenced by a variable not included in the experimental design, i.e., the anticipation of waiting for a test that is new to the testee. All measures should be taken at the same time of day, whether on a single subject or a number of subjects.

Performance test scores represent a quantity that the subject will produce rather than what he is capable of doing. Consistency of motivation, therefore, becomes a vital part of the experimental procedure. Anything from a loud noise to the presence of a friend or foe in the room provides an additional stimulus which may alter performance scores (6). Knowledge of his own results or preconceived notions of the desired outcome of the total investigation can alter a subject's response and change the validity of measures. Every effort must be made to keep the subject's total physiological effort constant whenever a quantity of work is to be imposed.

Repetition of performance tests leads to an improvement in efficiency of performance that can be attributed to learning and to training. To illustrate, several experimenters compared effects of a gelatin on performance, with varying degrees of improvement cited. None of them adequately controlled effects of training, even though Kaczmarek (16) indicated that his gains in performance are greater than training effects. Subsequently, Hellebrandt and others (12) controlled both diet and training and found no improvement that could be attributed to gelatin. Efficiency tends to vary with speed as well as with each individual. Although load may be adjusted so that metabolic costs are the same, the mechanical efficiency of any movement varies according to the speed at which it is performed (14).

EXPERIMENTAL PROBLEMS

Selection of a suitable problem for study is often so difficult for the inexperienced research worker that he must seek help from an expert adviser to avoid delay and floundering. Extensive reading in the area of research that is of interest to the student—plus the construction of an outline including the purpose, the method(s), the design, and the conclusions for each study reviewed—will help to give a background for the proposed study. Only then is he ready to attempt to outline in detail the problem of his choice. (See Chapter 3.)

In experimentation in the area of applied physiology, several questions must be answered before any study is undertaken: (a) Is the study feasible in terms of time and equipment available? (b) Is it possible to take the measurements that are desirable? (c) Are the researcher's skills and knowledges great enough to assure objective data? (d) Is it possible to control or equate all factors except the variable? (e) Is there a need and an application to the field for the proposed study? (f) Do more expert researchers consider the problem worthy? More time and effort will probably go into the planning of the study than into the actual collection of measurements. Often a preliminary study is desirable to try out the original ideas.

Three types of designs are popular in physiological research—(a) the single group with "pre" and "post" experimental variable (23, 27), (b) the matched or equated groups of which one may become a control group (22), (c) two or more selected groups which are subjected to two or more methods or tests that can be interpreted by variance analyses (32). In some few instances, a present physiological status needs to be determined (9, 15). Often the number of subjects used in physiological research is relatively small, in which case the number of determinations per subject is unusually large (2). It may be noted that experimental groups are rotated where practice or training effects may influence interpretation (27, 30).

BIBLIOGRAPHY ON APPLIED PHYSIOLOGY

1. Brouha, Lucien. "The Step Test: A Simple Method of Measuring Physical Fitness for Muscular Work in Young Men." *Research Quarterly* 14:31-36; March 1943.
2. Carlisle, Forbes. "Effect of Preliminary Passive Warming on Swimming Performance." *Research Quarterly* 27:143-51; May 1956.

3. COMROE, J. H. JR. "Pulmonary Function Tests." Section 2. (Edited by J. H. Comroe, Jr.) *Methods in Medical Research, Vol. 2.* Chicago: Year Book Publishers, 1952.
4. ELBEL, E. R., and GREEN, E. L. "Pulse Reaction to Performing Step-Up Exercises and Benches of Different Heights." *American Journal of Physiology* 145:521; 1946.
5. EYESTER, J. A. E. *The Clinical Aspects of Venous Pressure.* New York: The Macmillan Co., 1929.
6. FRASER, D. C. "The Relation of an Environmental Variable to Performance in a Prolonged Task." *Quarterly Journal of Experimental Psychology* 5:31-32; 1953.
7. GARRY, R.C.; PASSMORE, R.; WARNOCK, G. M.; and DURIN, J. V. G. A. "Studies on Expenditure of Energy and Consumption of Food by Miners and Clerks, Fife, Scotland." *Medical Research Council Special Report, Series No. 289.* London, 1955. p. 1-70.
8. GREULICH, W. W.; DAY, H. G.; LACHMAN, S. E.; WOLFE, J. B.; and SHUTTLEWORTH, F. W. *A Handbook of Methods for the Study of Adolescent Children.* Monograph of the Society for Research in Child Development, 3:2, Serial No. 15, 1938.
9. HALE, C. J. "Physiological Maturity of Little League Baseball Players." *Research Quarterly* 27:276-84; October 1956.
10. HAPP, W. P.; TUTTLE, W. W.; and WILSON, MARJORIE. "The Physiological Effect of Abdominal Cold Packs." *Research Quarterly* 20:153-61; May 1949.
11. HELLEBRANDT, F. A.; and MILES, M. M. "The Effect of Muscular Work and Competition on Gastric Acidity." *American Journal of Physiology* 102:258-66; 1932.
12. HELLEBRANDT, F. A.; RORK, R.; and BROGDAN, E. "Effect of Gelatin on Power of Women to Perform Maximal Anaerobic Work." *Proceedings Society Experimental Biology and Medicine* 43:629; 1940.
13. HENRY, F. M. "An Electrical Cardiometer." *Science* 86:2227-30; 1937.
14. HENRY, F. M.; DEMOOR, J.; and TRAFTON, I. R. "Individual Differences in Oxygen Metabolism of Work at Two Speeds of Movement." *Research Quarterly* 22:324-33; October 1951.
15. JOHNSON, B. L. "Influence of Puberal Development on Response to Motivated Exercise." *Research Quarterly* 27:182-93; May 1956.
16. KACZMAREK, R. L. "The Effect of Gelatin on the Work Output of Male Athletes and Non-Athletes and Girl Subjects." *Research Quarterly* 11:109-19; Dec. 1940.
17. KARPOVICH, PETER V. "A Frictional Bicycle Ergometer." *Research Quarterly* 21:210-15; October 1950.
18. KARPOVICH, P. V., and PESTROCOV, K. "Effect of Gelatin upon Muscular Work in Man." *American Journal of Physiology* 134:300; 1941.
19. KELSO, L. E. A., and HELLEBRANDT, F. A. "The Recording Electrodynamic Brake Bicycle Ergometer." *Journal of Laboratory and Clinical Medicine* 19:1105; 1934.
20. KEYS, A.; BROZEK, J.; HENSCHEL, A.; MICKELSEN, O.; and TAYLOR, H. *Human Starvation.* Vol. 2. Minneapolis: University of Minnesota Press, 1950.
21. KOFRANYI, E., and MICHAELIS, H. F. "Ein Tragborer Apparat zur Bestimmung des Gasstoffwechsels." *Arbeitsphysiologie* 11:148; 1941.
22. MICHAEL, E. D., JR. "Effect of Cold Showers on Circulation of Conditioned and Non-Conditioned Men." *Research Quarterly* 28:38-49; March 1957.
23. MICHAEL, E. D., JR., and HARDER, T. "Effect of Aqua Lung on Circulation of Adults While on Dry Land." *Research Quarterly* 27:403-408; December 1956.
24. MILLER, W. A., and ELBEL, E. R. "The Effect upon Pulse Rate of Various Cadences in the Step-Up Test." *Research Quarterly* 17:263; December 1946.
25. MOUTIS, N. P. "Cardiovascular Measurements on College-Age Smokers and Non-Smokers." *Research Quarterly* 26:454-60; December 1955.
26. ORSINI, D., and PASSMORE, R. "The Energy Expended Carrying Loads Up and Down Stairs Using the Kofranyi-Michaelis Calorimeter." *Journal of Physiology* 115:95; 1951.

27. PACHECO, B. A. "Improvement in Jumping Performance Due to Preliminary Exercise." *Research Quarterly* 28:55-63; March 1957.
28. PETERS, J. P., and VAN SLYKE, D. D. *Quantitative Clinical Chemistry.* Vol. 2. Methods. Second edition. Baltimore: Williams and Wilkins Co., 1932.
29. SALIT, ELIZABETH POWELL, and TUTTLE, W. W. "The Validation of Heart Rate and Blood Pressure Determinations as a Measure of Physical Fitness." *Research Quarterly* 15:252-57; October 1944.
30. SILLS, F. D., and O'RILEY, V. E. "Comparative Effects of Rest, Exercise, and Cold Spray upon Performance in Spot Running." *Research Quarterly* 27:217-19; May 1956.
31. SIMONSON, E.; SNOWDEN, A.; KEYS, A.; and BROZEK, J. "Measurement of Elastic Properties of Skeletal Muscle in Situ." *Journal of Applied Physiology* 1:512-25; 1948.
32. SKUBIC, V., and HODGKINS, J. "Effect of Warm-Up Activities on Speed, Strength, and Accuracy." *Research Quarterly* 28:147-52; May 1957.
33. SLATER-HAMMEL, A. T. "Influence of Order of Exercise Bouts upon Neuromuscular Tremor." *Research Quarterly* 26:88-95; March 1955.
34. TUTTLE, W. W. "The Use of the Pulse Ratio Test for Rating Physical Efficiency." *Research Quarterly* 2:5-17; May 1931.
35. TUTTLE, W. W.; JANNEY, C. D.; and THOMPSON, C. W. "Relation of Maximum Grip Strength to Grip Strength Endurance." *Journal of Applied Physiology* 2:663-70; 1950.
36. TUTTLE, W. W., and WENDLER, A. J. "The Construction, Calibration and Use of an Alternating Current Electrodynamic Brake Bicycle Ergometer." *Journal of Laboratory and Clinical Medicine* 30:173-83; 1945.
37. VAN HUSS, W. D.; MONTOYE, H. J.; REINEKE, E. P.; and COCKRELL, J. "Some Precautions in Exercise Metabolism Experiments." *Research Quarterly* 27:368-70; October 1956.
38. WEISS, R. A., and KARPOVICH, P. V. "Energy Cost of Exercise for Convalescents." *Archives of Physical Medicine* 28:447; 1947.

Applied Psychology

M. GLADYS SCOTT

Modern science has grown in the direction of numerous specializations. However, the boundaries between them are vague and overlapping. The academic disciplines are basically not as different as their respective members sometimes think. For example, the physiologist, psychologist, sociologist, anthropologist, and philosopher are all concerned with human behavior—the reasons why man behaves as he does and the bases on which his behavior is modified. The educator is also interested in the same problems. Skinner (67) ably summarizes the status of the experimental approach to behavior. He concludes with estimates of broad technological application and says that "the most exciting technological extension at the moment appears to be in the field of education."

Likewise, the physical educator shares these same interests. In fact, he probably needs to be conversant with more of the so-called disciplines, their research, and their methods of work than almost any of the members of those respective disciplines.

Because of this breadth of interest in human behavior, the physical educator needs to read from publications in many fields. From these diverse sources he will get the theories that contribute to basic understanding and to proposed research, new ideas for debate and experimentation, and ideas for integrating concepts and existing knowledge and bringing them to experimentation on technological applications. The reader cannot stop with *Psychological Abstracts*, though that is probably his starting point. (See Chapter 2 for suggestions on reading.) New psychology texts which will challenge and stimulate the physical educator's interest are important reading. Recent developments in fields such as neuro-anatomy, neuro-biology, neuropsychology must be followed carefully (49). They have implications for learning, response to emotional stress, motivation, psychotherapy, and other aspects of behavior of importance to the educator and to the research person (79). The guides of the experimental psychologist, such as Brown and Ghiselli (4) and Townsend (78), offer help in hypothesizing and designing studies, whether they are of a strictly psychological application or otherwise.

Psychological problems can be attacked by observational and descriptive investigations as well as by experimental methods. The nonexperimental methods aim to discover the behavior of individuals and differences between individuals, relationships which exist, and the adjustment of individuals to different conditions—and all this with an emphasis on cause. Where applicable, tests yield a better source of data than verbal reports from the subject on his experiences, sensations, or emotions. Though projective techniques have been developed, there is still the problem of dealing frequently with cases or groups of cases rather than with a sample representing a population. In fact, the population is not always clearly defined because of the nature of the traits or behavior studied.

LEARNING

Learning is a modification of response as a result of practice or experience. It is distinguished from sensory or affective response and from maturation. This does not mean that these may not all be taking place along with the learning process. However, research on learning must attempt to control these factors, or at least must recognize or use them in identifiable ways.

A basic problem of concern to the physical educator has had little recognition or at least comparatively little investigation. This problem is the relationship between acquisition of mental and motor learning. Studies on mental practice would lead one to believe that there may be common processes. However, the theories pertaining to reminiscence, empathy, retroactive inhibition, and facilitation have not been adequately tested on motor learning. Another problem of special concern to the physical educator is the relationship between the isolated, fine motor learnings of the psychology laboratory and the gross motor learnings of the physical education class. Similarities have been assumed, but complexities of tasks are known to affect the learning pattern. Certainly finger tapping, pursuit rotors, card sorting, typing, lever pulling, block placing, and pin sorting are different problems in complexity than a golf swing, a tennis serve, a dive, a trampoline or tumbling stunt, or a dance movement. However, familiarity with the findings on such studies is almost sure to yield some insight into the learning process, or into ways to set up learning experiments on other types of motor co-ordinations.

The psychologists attack the problem of learning from the dual approach of hypothesizing and experimentation. Their generalizations have resulted in several theories with respect to the learning process. Study of their comparative rationale and implications may be guided by such reading as McConnell (42), Thorpe and Schmuller (77), or Hilgard (30). The would-be experimenter or technologist is left, then, with the choice of following one particular theory or an eclectic route toward solution of his problem.

There is enough known about the factors influencing learning to provide bases for experimentation in physical education skills. Study of such variables requires them to be singled out and controlled. Ideally this should be in the real situation, i.e., gymnasium or playing field. The problems of control of human subjects and their environment have frequently led the psychologists to deal only with laboratory situations or with animals. Much of our knowledge of learning has been obtained from animals.

Laboratory studies employ equipment of some type (see Chapter 6) on which definite tasks are performed and on which objective records of proficiency can be obtained. These records may be successful acts per time unit; errors made per time unit, or per trial; or magnitude of errors made, as for example, total or average deviations made on an accuracy task. The same type of records can be used in the gross skills of physical education—for example, one might time the interval three balls could be juggled without error or the time to swim 100 yards. Errors are very obvious in bowling or archery and can be meaningful records of learning. Deviations from accuracy are frequently used in testing of kinesthetic functions, as in an approach shot in golf or placement of a tennis or badminton serve. Psychological studies can be invaluable in planning appropriate records of proficiency.

Learning records are usually plotted with practice units on one axis and proficiency units on the other. Learning curves reflect the magnitude and dissimilarity of proportions between the two axes. However, one of two principal types of curves is apt to be found. One type shows rapid acquisition of skill and a gradual or abrupt decrease in slope after a moderate amount of practice. The other type of curve is a double or "S" curve. Little or no improvement is made in the initial practice; then rapid learning occurs and follows the characteristics of the first type from that point on. The difference is apparently due to varying complexity of skill being practiced, to varying degrees of facilitation or inhibition from previous experience, or to the degree to which the records being taken indicate small increments of skill. Grant and co-workers (17) propose a possible statistical procedure for predicting later parts of a learning curve from any given part. Hayes (23) presented a technique for plotting of curves based on average performance with particular reference to learning in the immediate region of the criterion or near completion of learning. Rivlin (57) believes case studies of learning curves are more important than aggregate ones.

Plateaus and regressions are usually found in the learning curve. These are apparently not functions of the complexity of the skill but rather the elements to which the learner has to attend. When any new element is introduced, the slope of the learning curve decreases. For example, a change in range for the archery learner or addition of breathing to the co-ordinated crawl stroke causes apparent plateaus or temporary loss. Likewise, concentration on parts of a skill rather than on the whole seems to produce flattened curves. This type of information appears in psychological studies and is shown in learning curves which have been done in physical education.

Investigators of motor learning should consider carefully the shape of the learning curve and, if possible, determine at what point the leveling-off stage occurs, the probability of plateaus or troughs, and the amount of practice to get the learner beyond this level. Such predictions are not always possible, but failure to reckon with these characteristics of the curve can lead to failure of the experiment to show significant results. Many studies in physical education have failed because of this. Such failure is sometimes unavoidable because the length of the school term is not sufficient to get beyond this stage of learning. For example, in comparing two methods, both will probably show a plateau, but one may take the learner out of the plateau sooner than the other. If the experiment stops while both groups are on the plateau, the difference in the methods is not obtained.

Learning at different age levels may differ in rate and potential quality. The psychologists have worked mostly with young adults or children. These are the ages of most concern to the physical educator. However, if gerontology and recreation workers persuade the "oldsters" to acquire new recreational skills and hobbies, then this will become a fertile field for either the psychologist or physical educator.

One of the problems of determining learning is the index of accomplishment to be used. Numerical scores of increments of skill are not acquired with equal ease at all ability levels. This makes interpretation of learning curves difficult. In fact, the mere construction of a curve may be impossible because of the dissimilar nature of the co-ordination to be learned at different stages of learning. McCraw (44) summarizes some of the more common methods of trying to meet this difficulty and presents a comparison of several methods of measuring the learning for a group starting at different initial skill levels.

DeLong (9) summarizes recent trends in learning studies. These trends represent not only practice in research design but also attempts at solution of some of the problems of working with a behavioral trait as complex as learning. Reading of this report would be desirable for investigators in this area.

TRANSFER

Some of the earliest learning experiments in psychology dealt with the transfer of learning from one activity to the acquisition of the next. Perhaps the fact that the scientific curiosity of the psychologist had been partially satisfied on this matter long before physical education was doing much research may explain why the physical educator has not done much on this problem.

McGeoch (45), Stephens (73), and others provide a discussion of methods, experimental results and interpretation, and an extensive bibliography of studies on transfer. The student who wishes to work in this area will find it advantageous to start with this background. He may then proceed to the research reports, including those in physical education.

General principles and concepts would appear to be transferable. Motor learning offers a good opportunity to study transfer. Duncan (10) found that transfer increased directly with degree of inter-task-similarity. He attributed it to response generalization as well as learning how to learn. Barch (2) found that tasks of high difficulty had greater transfer than easy ones. Lindeburg (40) synthesized the findings and concluded a high level of specificity in the motor field, and therefore little transfer. Nelson (52) concluded from his data that skills involving similar elements and patterns should not be learned alternately, but rather consecutively. Deliberate teaching for transfer probably cannot change this.

Henry (27) showed that the improvement in reaction time acquired through negative motivation of electric shock is transferred from one skill to another of varying complexity within the limits of the laboratory experience. Munro (47) found that this improvement was retained for a period of six weeks and then began to regress. Fairclough (13), who participated in this same series of studies, worked on reactions of hand and foot. He concluded that transfer of training failed to occur but that transfer of motivated improvement did occur. This may help to clarify some of the difficulties encountered in trying to solve the matter of transfer. Practical experience with students has led many teachers to believe that some elements are transferable. In view of the developing knowledge of kinesthetic functioning, it is possible that new evidence may be uncovered to show means of improving transfer in motor learnings.

ATTENTION

Attention is a process of selective response to stimuli. The individual experiences at all times a number of stimuli, yet at any one moment he attends to only one, or at least very few, and disregards the rest. This selective process results in alertness with respect to part of the situation, a receptivity toward and vividness of experience inherent in the stimuli "attended," and a complete disregard and lack of cognizance of those "unattended."

Attention may then be considered a basis for perception, a stage of readiness or preparation as well as a continuing state of facilitating receptivity. The individual's physical state and needs are important determiners of attention. Environmental factors such as movement, bright colors, unusual objects, or action draw attention. These factors appear to be somewhat more important with children than with adults, leading one to assume that experience and learning lead to acquired determiners of a different type and an acquired inattention to many of the stimuli affecting the inexperienced and young. Movement stimuli, satisfactions from performing or moving, and acquired play interests all contribute toward attention of children to play situations within their level of comprehension. The attention of the youth or adult to play stimuli is apt to be only transitory with respect to the action itself, lacking with respect to physiological drive, and possible only if acquired interests are present or can be developed. These facts have real implications for both learning situations and curriculum planning.

An overt evidence of attention may be found in postural alertness and tension, immobility, fixed gaze, and sometimes even slow and shallow breathing. If the object of attention is stationary, the head and eyes seem to be fixed and even blinking is checked. If the object is in motion, the eyes follow it, as may be clearly seen by observing the audience at a tennis match. If there is no actual visual point attracting attention, this fixation of gaze is toward a point but with no real perception of that which is seen. This motor evidence of attention facilitates study of attention through muscle tension and visual processes.

The problems most frequently studied with respect to attention are span and range. These have usually been determined through visual or auditory stimuli and measured through motor response, steadiness, muscular tension, and respiratory response. Of specific interest for the physical educator is the study of attention span of children for experimentally designed toys (46).

Many of the methods of motivation or incentives are in the nature of attention determiners. Or they may be for development of interests to provide selective stimuli for continuing learning experiences. The positive and negative rewards used in the conditioning studies lead to the selective process of attention. This anticipatory state, coupled with comprehension, may account for much of the transfer of learning. Studies in this whole area of attention in relation to gross motor learning could yield significant results for the teaching of physical education activities and development of recreational interests and skills.

MOTIVATION

Experiments on the strength of physiological drives, usually hunger, have been made on animals rather than on human subjects for obvious reasons. But whether animal or human subjects are used, a meaningful incentive improves the rate of learning. The incentives used consist of appropriate forms of reward or punishment, praise or reproof, acceptance or rejection, or rivalry with reward at stake. The plan of providing punishment reveals that subjects respond equally well in their attempts to avoid the unpleasant. This is a negative approach which the educator is not apt to adopt deliberately, though it probably operates on students more frequently than is realized. Symonds (76) emphasizes that education should proceed on the basis of interest rather than fear of punishment. Wedemeyer (83) attributes "nonachievement" on the college level to the effect of unfavorable early experiences.

The investigations on human subjects includes all ages from the very young child to the adult. The studies of Lewin and co-workers (39) are concerned with changing needs of the child during his development and with the effect these changes have on behavior. Their studies on level of aspiration and the effects of success and failure are directed toward understanding of the emergence of the personality pattern and the subject's recognition of "self." Havighurst (22) presents case studies exemplifying this type of influence on behavior. This is an area where potentials in physical education are great because success can be clearly demonstrated and often measured. It is a success in which the child has a compelling interest. It is an area where rivalry for recognition often brings reward (team membership) for the few and punishment (deprivation of team participation, and often of other activity, too) for the many. Studies like that of Smith (72) have pursued this particular problem, but they are really few in number.

Animal experimentation relative to neuro-anatomy, motivation, and learning have led to some interesting observations. Olds (50, 51) has located what he calls *pleasure centers* in the brain and shows that the electrical stimulation of these centers produces results comparable with those of reward. Other neuro-anatomists have found similar reward effects which are apparently effective motivation for learning. Further study from such sources should help us to understand how motivation operates in the more complicated human process of values and meaning in relation to neuromuscular acquisitions.

Pfeiffer (53), writing extensively on functioning of the brain, says, "There is nothing so tenaciously inborn in us, no process so deep-rooted, that we cannot modify it appreciably—providing we have good reasons to do so."

How an incentive operates to facilitate human learning is not clear. It is probably through various effects. Recognition of a positive or negative reward in the initial stage may gain attention. It is frequently a more tangible and meaningful goal than the acquisition of the learning itself. During learning it tends to focus effort and to minimize distractions. It frequently has associations or qualities arousing affective response and thereby enlists the extra effort characteristic of emotional release. The reward may be more nearly the common element from one learning situation to another and thereby lay the foundation for a transfer of learning. Certainly, an incentive which would do all these things would be a tremendous aid throughout the learning process.

Motor learning has one advantage—that of results being fairly obvious in most instances. Knowledge of results has long been believed to be an incentive. Practice tests or more formal achievement tests have been used on the premise that knowledge of results is a motivator. Howell (31) demonstrated that knowledge of the time-force factor of a racing start improved learning rate above that for a control group which did not have this information. Henry (27, 28) applied electric shock when reaction was slow. He found it gave significant improvement in reaction time but concludes that this is due to "the informative value of the motivating stimuli rather than to punishment as such or to a direct facilitative function" (28).

On the other hand, Johnson (36) found that effort under incentives of competition and direct verbal encouragement and exhortation resulted in increased work in some cases, in decreased amounts in others. Apparently adverse physiological conditions are more apt to follow, as evidenced by nausea. This reaction has been found in other studies.

EMOTIONS

Both the physiologists and psychologists have studied the effects of emotions on body functions. The work of Cannon (7) is a classic in this field. One of the problems in such study is the difficulty of isolating a given emotion, or evoking a specific intensity in any one subject or comparable intensity in all subjects. Those studied have been enthusiasm vs. irritation, joy vs. anger, satisfaction vs. annoyance, anticipation vs. fear. Perhaps these could be more easily produced in the physical education setting than in that of the laboratory. These are all important to the physical educator. Fear is especially so and its development has been discussed by Kingsley (37). Physical education studies on overcoming fear appear sporadically, but their number is encouraging.

The studies of neural structure and function lead toward evidence of existence of circuits in which cortical activity can dominate emotions as well as behavior. This would lead to possible implications in relation to mental health (79). Pribram and Kruger (56) make an excellent approach to synthesizing implications of many studies and to the process of hypothesizing and planning further exploration into a new frontier of knowledge.

The techniques for measuring emotional response developed by the physiologists and psychologists include changes in respiration, circulation and blood pressure, galvanic skin response, muscular tension, and overt expressions of face or voice. The latter are least reliable, and are useful primarily for work with infants. Few of these have been used successfully with stimuli from physical education situations. Some examples can be found (68, 84) and would appear to suggest a promising field for further study.

MUSCULAR TENSION

The physical education teacher is apt to think of muscular tension in terms of certain specific states. The coach knows the tension of anticipation which mounts if activity is impossible. The teacher knows the rigidity, almost a spastic state, which results when the beginner is determined to learn or when he reacts to frustrations of errors by fixation of purpose and ever increasing effort. Likewise, the teacher and the health counselor know the chronic hypertension of the person who lives and works on an emotional level in which he frequently feels that solution of his problems is impossible. The recreation leader sees the person who says, "I want to get some exercise or hobby so I can relax." This knowledge has led to lack of clarity in dealing with the condition in the classroom and to lack of adequate research geared to problems of the physical and health educator.

There appear to be two kinds of tension—that which inhibits and that which facilitates action of the individual. The first is apparently due to an emotional stress and the second is a concomitant of effort. This was pointed out long ago by Bills (3).

Measuring devices include those indicating reflex responses, resistance to movement, range of motion, tremor, and electrical responses of the muscle. (See Chapter 6.) Those interested in relaxation have dealt primarily with the latter of these; those interested in problems of movement have dealt with the first named.

Tension is a practical problem which needs careful study. In the underlying premise of studies in this area, it should be recognized that the inhibitive and facilitative condition of tension may be of different magnitude, may result in different kinds of overt response, and may respond to different stimuli. Only careful research will answer these points.

SENSATIONS

Perception is an accompaniment of sensory experience. Interpretation or meaning attributed to a perception serves to give us awareness and knowledge of the world, our environment, and ourselves as an organism in that world. The sensations which provide these experiences are vision, hearing, smell, taste, touch, and kinesthesis. The first three deal with stimuli completely removed from contact with the organism. The object seen, heard, or smelled remains removed from the organism unless the perception and meaning cause the person to touch it, taste it, or otherwise react with respect to it. Taste and touch are dependent upon contact with the source of stimuli. In contrast, kinesthetic stimuli originate within the organism.

The common elements for all these sensations are special sensory organs, special nerve endings as receptors, special stimuli, and specificity of reception and perception. All have a psychological-cultural aspect. This is clearly presented in a discussion of the psychophysiology of taste (80). The health educator might find some value here relative to food habits and their development.

Vision. Visual sensation and perception involve highly complicated processes. In considering any population of subjects, certain variations in visual acuity will be found. Kingsley and Garry (37) estmate that among elementary school children 10 percent or more have serious defects and another 10 percent have minor disabilities which can impair response to learning situations. Dalton (8) found that less than 20 percent of 5,000 elementary school children studied could pass all tests given with the Keystone Telebinocular. This finding not only arouses concern for the classroom situation and learning in general, but also raises some real problems for various kinds of research.

Binocular vision is basic to depth perception. The two eyes see slightly different views of close objects; the brain interprets the two bits of information and attempts an objective synthesis. This process can be studied conveniently by the stereoscope. Objects at too great a distance appear too much alike on the retinal images, and depth perception is lacking. Studies on crucial distances in this respect and in individual differences might solve certain problems of safety, and lead to an understanding of variation of visual perception, as well as accommodation to changing size of courts and variations in lighting on the athletic field.

Retinal rivalry and one-eyed dominance sometimes result from inability of the neural centers to synthesize the two images. Dominance in use of the eyes may also be due to actual differences of vision. Whether habit is a factor does not appear clear. It has been demonstrated that eye dominance exists and can be measured (21).

The size of the visual field can be measured with the compimeter (41). Slater-Hammel (71) showed increased reaction time with increased range of stimuli into the peripheral field. It appears that the peripheral range may be increased with practice. The exact explanation is not clear. It may be that practice facilitates awareness of stimuli in the periphery, or delays fatigue, or perhaps trains the muscles of the eye so as to adapt eye position to the conflicting need of focus and peripheral stimulation. The importance of peripheral vision to sports and athletics would indicate a need for further research in this area.

Intermittent light at slow, regular intervals is perceived as a flicker. Increased frequency eliminates the flicker effect and is perceived as motion, if variations of light or images are involved, or as constant light, if only uniform light is involved. The frequency at which flicker disappears is known as the critical fusion frequency. This critical frequency varies with the excitability of the visual-cortical centers. Henry (24) described a device for measurement of critical fusion frequency. He found that physiological states apparently affect it. For example, hard physical work depresses it and light exercise and cold hip baths increase it. Further study is indicated. Simonson (65) proposed that this critical point could be used as a measure of fatigue of the central nervous system.

Most sports involve reaction to a moving ball. The usual coaching cue is "keep your eye on the ball," and it is assumed that it is possible to do so up to the instant of catching or hitting. Hubbard and Seng (32) studied this possibility in batters in professional baseball and cast doubt on such continuous focus. Clarification of this might help both the performer and the coach. A related problem is that of the blackout period during a blink. This was studied by Slater-Hammel (69).

Brown (5) suggested a possibility of use of visual after-image. He applied it to the learning of golf.

There is need for the technological application to physical education and athletics of much of the understanding of visual perception. In addition to those indicated above, there would be such problems as color vision, color blindness, and color relative to peripheral vision. There are also the problems of visual accommodation, relationship to neuromuscular control, and variations under different physiological states, effects of smoking, and perhaps others pertinent to sports and athletic participation.

Hearing. The problems of hearing are somewhat similar to those of vision. Individual differences exist which are too frequently ignored, even though techniques of measuring acuity and diagnosing type of defect have been greatly improved.

Little research has been done on the importance of auditory cues in sports or athletic situations. Experience has taught the difference in response of subjects to sharp, incisive sounds in contrast to muted or vague commands. Localization of source of sound, or its direction, is a function of binaural hearing, aided by turning of the head. Right-left differentiation is much more efficient than that in another direction. Yet observation of reactions of the nonsighted performer shows us that auditory stimuli can be used, and apparently perception developed much beyond what the majority of sighted subjects ever experience.

Kinesthesis. Kinesthesis is sometimes referred to as h sense. It is the sense by which the person is aware of n of body parts, of their movement, rate, and range, and of otal body movement and position. Probably it is more complex in its source of pertinent stimuli in most of our daily experiences than any of the senses. It is the only one of the sensations in which the actual source of the stimulation is inherent within the structural and functioning mechanism of the body. The special receptors are located in the muscles, tendons, ligaments, and articular cartilages, and in the epithelial lining of the semicircular canals of the inner ear. Information so obtained is supplemented by visual cues. All are synthesized in the central nervous system and the response is sometimes a reflex control, sometimes a link in the chain of responses in a highly co-ordinated skill, and sometimes merely an awareness of perception of movement or position which the person may describe verbally, control motorly, or voluntarily depart from with precision to another position, at the same or a different tempo.

There is frequently difficulty in separating the more truly kinesthetic sensation and perception from that of touch or the tactile sense, and even that of vision. This has partially contributed to the difficulty of measuring kinesthetic acuity in the sense that one measures vision or hearing. Work on this problem to date seems to indicate a very high degree of specificity of the elements of kinesthesis or the expressions of kinesthesis. Developments in the measurement of kinesthesis have come through efforts of the physical educator rather than the physiologist or psychologist. These efforts can be traced through studies such as Young (91), Phillips (55), Seashore (64), Russell (61), Witte (87), Roloff (58), Stevens (74), Wiebe (86), and Scott (62). Much more needs to be accomplished in refinement of measures and determination of the most representative sampling of the components of kinesthetic sensation. The basis will then be laid to determine the possibilities for training of acuity of perception, the relationship with facility of learning motor skills, and perhaps improvement of method of teaching by better understanding of the use of imitation, pace of empathy in learning, or discovery of the common element on which transfer of training can be achieved. These are primarily problems for the physical educator to deal with, since they affect his efficiency of teaching.

Phillips and Summers (55) made a contribution toward the clarification of the question of kinesthesis and motor performance. They found some relationship with learning of bowling—higher at beginning stages than with more skilled bowlers. They also found significant differences in kinesthetic perceptivity between the preferred and nonpreferred arms. Investigations in this area might lead to diagnostic measures relative to the various physical education activities.

Equilibrium and balance are two expressions of kinesthetic response which have been studied extensively. The various types of apparatus (see Chapter 6) are used primarily to measure sway or steadiness of position. This is important relative to problems of posture. However, static balance appears to have little relationship to dynamic balance. It is not clear what relationship may exist to gross motor skills in general. Espenschade and co-workers (11) did find differences in dynamic balance between athletes and nonathletes. Estep (12) found static balance, as judged by body sway, to be related significantly to both sports and rhythmic abilities. Both static and dynamic balance appear as distinct elements in all studies of kinesthesis to date.

Rotatory action seems to have some bearing on kinesthesis, or at least on the two aspects of balance. Studies to date are very inconclusive, but the more marked physiological effects in the form of motion sickness point toward possible application in physical education activities as well as toward better understanding of the process of kinesthetic stimulation. Much work needs to be done on this phase.

Physical educators frequently refer to eye-hand co-ordination. This includes skills such as catching, throwing, striking. The psychologist approaches this through tests such as pursuit pendulum, pursuit rotor, and target throwing. In general, most of these appear to be highly specific. Investigations on the relationship of kinesthetic acuity and training in developing kinesthetic awareness might yield valuable information for the physical education teacher.

Orientation in space seems to be one aspect of kinesthesis (89, 90). It is a sense developed deliberately in the training of the blind person. The degree to which this and other aspects of kinesthetic perception can be trained to greater levels of acuity seems to point toward potential in the motor training of all persons. The inability to make discriminations in smaller ranges characterizes the average subject. For example, Phillips (54) and Young (91) both found differentiation in weights to be an unreliable measure. Perhaps it is complicated by other factors and might be considered in much the same light as other forms of perception. For example, differentiation in length of time intervals is poor, or at least markedly influenced by the nature and number of experiences occurring during the interval. Likewise, lines and figures are perceived differently in different settings and viewings by the same subject. Henry (26) found an increment of 1.25 pounds required to perceive a change in pressure, though actual error in maintaining uniform pressure was only .71 pounds. Slater-Hammel (70) studied this consistency of effort by means of the muscle potential measured on an electronic voltmeter. It has been thought that physiological states, particularly fatigue, might influence the reactions. For example, fatigue is assumed to make for inco-ordinations or motor errors, to detract from correct judgment of passage of time in a game, to create false feelings of heaviness in body segments, and to inhibit steadiness. Studies to date do not substantiate this assumption. Insufficient work has been done to indicate the effect of practice on successive trials in kinesthetic measures.

RHYTHM

Rhythm is a regular recurrence of patterns, successive stimuli, or impressions. It may be perceived through visual, auditory, or kinesthetic channels. Tests of time, intensity, and rhythmic pattern such as those of Seashore (63) are well known. However, ability on such tests does not correlate highly with dance ability or other rhythmic motor response. This may be explained by the fact that the subject taking a Seashore test makes a cognitive response and not a rhythmic one. It seems very possible that rhythmic ability is closely associated with kinesthetic ability, or at least certain aspects of kinesthesis. It might be assumed that rhythmic accuracy is dependent upon ability to perceive time intervals and adjust tempo and range of movement appropriately, that is, to make movements of the body or its segments in prescribed positions, planes, and tempo, and frequently in relation to auditory or visual cues.

Research in physical education has concentrated upon the motor response of the subject to an auditory rhythmic pattern as seen in Buck (6), Mussey (48), and Haight (18). Apparatus has not been very practical, and responses have tended to consist of finer movements rather than gross body movements characteristic of dance. Again, relationship to dance performance is poor. Ashton (1) presented a test of basic rhythmic patterns based on subjective ratings. This is subject to all the problems of the use of subjective ratings but would appear to be a possibility for further study. Waglow (81) developed a somewhat similar device for social dance.

Motor skills always involve a temporal sequence. Rhythm can be used in establishing the correct temporal relationships of co-ordination. Music is often used to set the tempo of work or dance movement. Since rhythm is conducive to relaxation, it is often used to avoid fatigue.

To secure maximum benefits from rhythm in the learning process, it must be used at the right time. The basic pattern of the movement must have been learned, i.e., movement in correct form and sequence. On the other hand, if used too late, it must be individualized in tempo or the individual must be left to practice at his own best speed as noted in Kingsley (37). Music has been used as a teaching aid by many teachers, but relatively few studies have been carried out to determine optimum usage.

EMPATHY

Empathy is a projection of self into, or identification of self with, an object, person, or action. It is probably the least understood of the attitude-emotional responses of the child or adult. Certain things appear to be essential for empathetic experience. First, the subject must have the opportunity and ability to observe the situations or persons with whom this interaction is to take place. Secondly, the subject must have previous experience which makes for accurate perception of what he observes. Thirdly, the subject must respond freely and without distractions, for the time being at least.

Empathy has long been recognized as a basis for enjoyment of art, of the beauty and grandeur of nature. Certainly the same can be said for appreciation of quality in movement; for dance as an art form; for impressions given by body postures, gestures, and other movements. Perhaps the same kind of experiences explain at least in part the enjoyment of play, of spectator sports, of satisfactions of teaching the beginner a motor skill. Even the personal interaction within a group may be affected by empathy (59). These are hypotheses which might well be tested by careful research.

Probably the kinesthetic sensation is important in empathy. Experience indicates a very different empathetic experience in the observation of known skills in contrast to that of unfamiliar ones, when previous experience in the skill has occurred, and when the subject attends mentally and with changing muscular tonus appropriate to the sequence of parts of the motor complex. Here is another possible extension of the use of kinesthesis in everyday experiences, of improvement of learning through better use of demonstrations, of better appreciation of movement as observed in any situation, and of training for appreciation of skill and artistic qualities of dance or sports.

DOMINANCE

The physical education teacher, like the teacher of any other manipulatory skill, is well aware of the variations exhibited by most subjects in facility and preference for using right or left hand. Psychologists have debated over the neural explanation for these variations, while the educational psychologists have debated the advantages and disadvantages of imposing uniform work habits.

The first interpretation of dominance assigned rather obvious categories—the right- and left-handed. There are, of course, variations of these depending upon the habits developed. From the manipulative standpoint, there are those who use either hand readily and successfully and a few who have no real preference.

The matter of dominance is broader than handedness. It includes feet and eyes, and Sinclair and Smith (66) suggest also dominance in the side on which breathing is done in swimming.

Dominance is determined by tests as described by Hildreth (29). These are so arranged that practice effects cannot cover up natural tendencies. The preferred hand is the one the subject considers dominant and is not always the one determined by dominance tests.

In physical education the usual practice is to permit the student to learn all skills as either a left-hander or a right-hander unless equipment, such as golf clubs or hockey sticks, restricts. Comparatively little has been done to study learning problems in relation to gross motor skills and dominance factors. Fox (14) concluded that the beginning bowler should use the preferred hand rather than the dominant one. This is doubtless due to habit developed in other skills. Sinclair and Smith (66) found that "laterality" in breathing was a more important factor in determining side and crawl stroke patterns than was either eye, hand, or foot dominance. Their summary fairly well covers the present situation with respect to this problem. They point out the complexity of laterality as a factor in motor learning, recommending that the teacher should expect and promote a consistent pattern and that attention to movement patterns as related to laterality should be given, particularly at the elementary level.

PLAY

Play is one of several forms of activity used in the processes of education. In the use of play as an educational tool, it is important for all teachers to understand the motives of the individual playing and the characteristics of play at the successive age levels. For the physical education teacher it is doubly important that the psychological implications of play activities be understood. However, relatively little has been done to study the actual motives and attitudes of the child at play.

In setting up a research study in this area, it is imperative to understand the theories of play and to establish clear-cut hypotheses relative to the psychological function of play. A helpful summary can be found in Wheat (85:41). New developments such as sociometry and the sociodrama provide new tools to be used. (See Chapter 5.) Studies in this field are interrelated with those of attitudes, motivation, and personality.

PERSONALITY

Personality is defined in various ways which reflect the thinking of the writer relative to the organization of the individual's behavior. Certainly, an individual's behavior is affected by his attitudes toward people, things, and circumstances. It is also affected by his traits or manner of responding—an intimate, personal, or unique organization of these traits. It is the organization of the traits in relation to an individual's needs, drives, and ability—and the relative strength of expression of certain traits—that makes the personality. An investigator in the area of personality must have a clear-cut operational definition as a guide to formulation of his study. The importance of the concept as a working basis is brought out by authors such as Stephens (73).

Since attitudes represent one phase of the personality structure, consideration should be given to methods of determining attitudes. This can be done by a direct approach with attitude scales or rating scales, man-to-man scales, etc. (See Chapter 5.) It may also be done by indirect methods.

These indirect methods vary in use and intent. In dealing with groups, particularly if trying to modify attitudes, the sociodrama may be used. In dealing with individual assessment, other techniques would be more suitable and would aim at understanding the total personality rather than the respective traits. These are sometimes classed collectively as projective techniques. They include personality inventories, the Thematic Apperception Test, and ink blot tests—each in several variations. (See Chapter 5.) Hurley (33) gives one of the newer versions of the TAT. Summaries of the application of these techniques can be found in reports by Furst (16) and Rothney (60). There has been some use made of these in physical education, with special designs to fit the situation and with free or unstructured response. They appear to have promise in revealing general patterns of response, emotional stress in relation to certain stimuli, and individual differences in reaction.

A newer development in this respect, growing out of social and personal status of the individual, is the "Who Am I?" type of test (38). It appears to be indicative of the individual's concept of self and his role in his environment. There is also some evidence that it may differentiate between professional groups. It has been used with physical education teachers and students (34).

Since personality is usually interpreted as having certain social implications, special techniques have been developed to determine social interaction, friendships, popularity, leadership status, etc.

The sociometric techniques (35) have been used enough in physical education situations to establish their worth and to indicate their value as measures of social and emotional learnings (82, 93). Hale (19) proposed criteria for better interpretation of the sociogram and comparison of groups at different times, thus making the measurement of changes more feasible.

There would appear to be a promising field for the physical educator in the study of personality. It is a challenge which could lead to better understanding of the individual, of class groups, of social growth of individuals, and of maturity of adjustment.

BIBLIOGRAPHY ON APPLIED PSYCHOLOGY

1. ASHTON, DUDLEY. "A Gross Motor Rhythm Test." *Research Quarterly* 24:253; October 1953.
2. BARCH, ABRAM M. "The Effect of Difficulty of Task on Proactive Facilitation and Interference." *Journal of Experimental Psychology* 46:37; July 1953.
3. BILLS, A. G. "Facilitation and Inhibition in Mental Work." *Psychological Bulletin* 34:286; May 1937.
4. BROWN, CLARENCE W., and GHISELLI, EDWIN E. *Scientific Method in Psychology* New York: McGraw-Hill Book Co., 1955. 368 p.
5. BROWN, DONALD W. "Seeing in the Dark." *The American Teacher* 34:11; December 1949.
6. BUCK, NADINE. "A Comparison of Two Methods of Testing Response to Auditory Rhythms." *Research Quarterly* 7:38; October 1936.
7. CANNON, W. B. *The Wisdom of the Body*. New York: W. W. Norton and Co., 1939.
8. DALTON, M. M. "A Visual Survey of 5000 School Children." *Journal of Educational Research* 37:81; 1943.
9. DELONG, ARTHUR R. "Learning." *Review of Educational Research* 25:438; December 1955.
10. DUNCAN, CARL P. "Transfer in Motor Learning." *Journal of Experimental Psychology* 45:1; January 1953.
11. ESPENSCHADE, ANNA; DABLE, ROBERT R.; and SCHOENDUBE, ROBERT. "Dynamic Balance in Adolescent Boys." *Research Quarterly* 24:270; October 1953.
12. ESTEP, DOROTHY P. "Relationship of Static Equilibrium to Ability in Motor Activities." *Research Quarterly* 28:5; March 1957.
13. FAIRCLOUGH, R. H., JR. "Transfer of Motivated Improvement in Speed of Reaction and Movement." *Research Quarterly* 23:20; March 1952.
14. FOX, MARGARET G. "Lateral Dominance in the Teaching of Bowling." *Research Quarterly* 28:327; December 1957.
15. FRANDSEN, ARDEN N. *How Children Learn*. New York: McGraw-Hill Book Co., 1957. 546 p.
16. FURST, EDWARD J., and ERICKE, BENNO G. "Development and Applications of Structured Tests of Personality." *Review of Educational Research* 26:26; February 1956.
17. GRANT, DAVID A.; KAESTNER, NOEL F.; and SCHIPPER, LOWELL M. "Autocorrelation Analysis of Gross Learning Scores." *Perceptual and Motor Skills* 5:53; June 1955.
18. HAIGHT, EDITH C. "Individual Differences in Motor Adaptation to Rhythmic Stimuli." *Research Quarterly* 15:38; March 1944.

19. HALE, PATRICIA WHITAKER. "Proposed Method of Analyzing Sociometric Data." *Research Quarterly* 27:152; May 1956.
20. HARMAN, J. M., and MILLER, A. G. "Time Patterns in Motor Learning." *Research Quarterly* 21:182; October 1950.
21. HART, DOROTHY MAE. "Factors Which Contribute to Success in Target Archery." Doctor's thesis. Iowa City: State University of Iowa, 1955. (Microcard PE 246.)
22. HAVIGHURST, ROBERT J. *Human Development and Education.* New York: Longmans, Green & Co., 1953.
23. HAYES, KEITH J. "The Backward Curve. A Method for the Study of Learning." *Psychological Review* 60:269; July 1953.
24. HENRY, FRANKLIN. "An Electronic Apparatus for Testing Fatigue by the Visual Flicker Method." *Journal of Experimental Psychology* 31:538; December 1942.
25. HENRY, FRANKLIN. "Discrimination of the Duration of a Sound." *Journal of Experimental Psychology* 38:734; December 1948.
26. HENRY, FRANKLIN. "Dynamic Kinesthetic Perception and Adjustment." *Research Quarterly* 24:176; May 1953.
27. HENRY, FRANKLIN. "Increase in Speed of Movement by Motivation and Transfer of Motivated Improvement." *Research Quarterly* 22:219; May 1951.
28. HENRY, FRANKLIN. "Independence of Reaction and Movement Times and Equivalence of Sensory Motivators of Faster Response." *Research Quarterly* 23:43; March 1952.
29. HILDRETH, GERTRUDE. "Development and Training of Hand Dominance, Characteristics of Handedness." *Journal of Genetic Psychology* 75:197; December 1949.
30. HILGARD, ERNEST. *Theories of Learning.* New York: Appleton-Century-Crofts Co., 1948.
31. HOWELL, MAXWELL L. "Use of Force-Time Graphs for Performance Analysis in Facilitating Motor Learning." *Research Quarterly* 27:12; March 1956.
32. HUBBARD, ALFRED W., and SENG, CHARLES N. "Visual Movements of Batters." *Research Quarterly* 25:42; March 1954.
33. HURLEY, J. R. "Iowa Picture Interpretation Test, A Multiple Choice Variation of the Thematic Apperception Test." *Journal of Consulting Psychology* 19:372; 1955.
34. ISENBERGER, WILMA E. "Self-Attitudes of Women Physical Education Majors as Related to Measures of Interest and Success." *Research Quarterly* 30:167; May 1959.
35. JENNINGS, HELEN H. *Sociometry in Group Relations.* Washington, D.C.: American Council on Education, 1948.
36. JOHNSON, BERGER L. "Influence of Puberal Development on Responses to Motivated Exercise." *Research Quarterly* 27:182; May 1956.
37. KINGSLEY, HOWARD L. *The Nature and Conditions of Learning,* Second edition. (Revised by Ralph Garry.) Englewood Cliffs, N. J.: Prentice-Hall, 1957. 565 p.
38. KUHN, MANFORD, and MCPARTLAND, THOMAS S. "An Empirical Investigation of Self-Attitudes." *American Sociological Review* 19:398, 1954.
39. LEWIN, K.; DEMBO, T.; FESTINGER, L.; and SEARS, P. "Level of Aspiration." *Handbook of Personality and the Behavior Disorders.* (Edited by J. McVicker Hunt.) New York: The Ronald Press Co., 1944. 296 p.
40. LINDEBURG, F. A. "A Study of the Degree of Transfer Between Quickening Exercises and Other Co-ordinated Movements." *Research Quarterly* 20:180; May 1949.
41. MCCLURE, J. A. "The Development and Standardization of a New Type Test of Peripheral Vision." *Journal of Applied Psychology* 30:340; August 1946.
42. MCCONNELL, T. R. "Reconciliation of Learning Theories." *Psychology of Learning.* Forty-First Yearbook, Part II. National Society for Study of Education Chicago: University of Chicago Press, 1942. Chapter 8.

43. McCraw, L. W. "Comparative Analysis of Methods of Scoring Tests of Motor Learning." *Research Quarterly* 26:440; December 1955.
44. Mc w, L. W. "A Comparison of Methods of Measuring Improvement." *Research Quarterly* 22:191; May 1951.
45. McGeoch, J. A. *The Psychology of Human Learning.* New York: Longmans, Green & Co., 1946.
46. Moyer, Kenneth E., and Gilmer, B. von Haller. "Attention Span of Children for Experimentally Designed Toys." *Journal of Genetic Psychology* 87:187; December 1955.
47. Munro, Sanford J. "The Retention of the Increase in Speed of Movement Transferred from a Motivated Simpler Response." *Research Quarterly* 22:229; May 1951.
48. Muzzey, Dorothy. "Group Progress of White and Colored Children in Learning a Rhythm Pattern." *Research Quarterly* 4:62; October 1933.
49. National Research Council, Survey of Neurobiology, Publication 237. Washington, D. C.: National Academy of Science, 1952.
50. Olds, James. "Pleasure Centers in the Brain." *Scientific American* 175:105; October 1956.
51. Olds, James, and Milner, Peter. "Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brain." *Journal of Comparative and Physiological Psychology* 47:419; December 1954.
52. Nelson, Dale O. "Studies of Transfer of Learning in Gross Motor Skills." *Research Quarterly* 28:364; December 1957.
53. Pfeiffer, John. *The Human Brain.* New York: Harper & Brothers, 1955.
54. Phillips, B. E. "The Relationship Between Certain Phases of Kinesthesis and Performance During the Early Stages of Acquiring Two Perceptuo-Motor Skills." *Research Quarterly* 12:571; October 1941.
55. Phillips, Marjorie, and Summers, Dean. "Relation of Kinesthetic Perception to Motor Learning." *Research Quarterly* 25:456; December 1954.
56. Pribram, Karl H., and Kruger, Lawrence. "Functions of the 'Olfactory Brain.'" *Annals of the New York Academy of Sciences* 58:109; March 1954.
57. Rivlin, Harry N. "Psychology of Learning." *Review of Educational Research* 26:246; June 1956.
58. Roloff, Louise L. "Kinesthesis in Relation to the Learning of Selected Motor Skills." *Research Quarterly* 24:210; May 1953.
59. Rose, Grace; Frankel, Norman; and Kerr, Willard. "Empathetic and Sociometric Status Among Young Teenagers." *Journal of Genetic Psychology* 89:277; December 1956.
60. Rothney, John W., and Heiman, Robert A. "Development and Application of Projective Technics." *Review of Educational Research* 26:56; February 1956.
61. Russell, Ruth I. "A Factor Analysis of the Components of Kinesthesis." Doctor's thesis. Iowa City: State University of Iowa, 1955. (Microcard PH 36.)
62. Scott, M. Gladys. "Measurement of Kinesthesis." *Research Quarterly* 26:324; October 1955.
63. Seashore, Carl. *Measures of Musical Talent.* Camden, N. J.: Educational Department of RCA Manufacturing Co.
64. Seashore, H. G. "The Development of a Beam-Walking Test and Its Use in Measuring Development of Balance in Children. *Research Quarterly* 18:246; December 1947.
65. Simonson, Ernst; Enzer, Norbert; and Benton, Roy A. "The Influence of Muscular Work and Fatigue on the State of the Central Nervous System." *Journal of Laboratory and Clinical Medicine* 28:1555; October 1943.
66. Sinclair, Caroline, and Smith, Inez M. "Laterality in Swimming and Its Relationship to Dominance of Hand, Eye, and Foot." *Research Quarterly* 28:395; December 1957.
67. Skinner, B. F. "The Experimental Analysis of Behavior." *American Scientist* 45:343; September 1957.

68. SKUBIC, ELVERA. "Emotional Responses of Boys to Little League and Middle League Competitive Baseball." *Research Quarterly* 26:342; October 1955.
69. SLATER-HAMMEL, A. T. "Blackout Interval During Eye Blinks." *Research Quarterly* 24:362; October 1953.
70. SLATER-HAMMEL, A. T. "Measurement of Kinesthetic Perception of Muscular Force with Muscle Potential Changes." *Research Quarterly* 28:153; May 1957.
71. SLATER-HAMMEL, A. T. "Reaction Time to Light Stimuli in the Peripheral Visual Field." *Research Quarterly* 26:82; March 1955.
72. SMITH, C. H. "Influence of Athletic Success and Failure on the Level of Aspiration." *Research Quarterly* 20:196; May 1949.
73. STEPHENS, J. M. *Educational Psychology.* New York: Henry Holt & Co., 1956, 690 p.
74. STEVENS, MILDRED. *Measurement of Kinesthesis on College Women.* Doctor's thesis. Bloomington: Indiana University, 1950. (Microcard PE 93.)
75. STEVENS, S. S., and DAVIS, H. *Hearing.* New York: John Wiley & Sons, 1938. 472 p.
76. SYMONDS, P. M. *Diagnosing Personality and Conduct.* New York: The Century Co., 1931. 572 p.
77. THORPE, LOUIS P., and SCHMULLER, ALLEN M. *Contemporary Theories of Psychology.* New York: Henry Holt & Co., 1954. 480 p.
78. TOWNSEND, JOHN C. *Introduction to Experimental Method for Psychology and the Social Sciences.* New York: McGraw-Hill Book Co., 1953. 220 p.
79. UPJOHN COMPANY. "Physiology and Pharmacology of Emotion." *Scope* 4; Winter 1955.
80. UPJOHN COMPANY. "The Psychophysiology of Taste." *Scope* 4; Winter 1955.
81. WACLOW, I. F. "An Experiment in Social Dance Testing." *Research Quarterly* 24:102; March 1953.
82. WALTERS, C. ETTA. "A Sociometric Study of Motivated and Non-Motivated Bowling Groups." *Research Quarterly* 26:107; March 1955.
83. WEDEMEYER, CHARLES A. "Gifted Achievers and Non-Achievers." *Journal of Higher Education* 24, 25; January 1953.
84. WELLS, PHILIP V. "Emotion in Fitness Tests." *Research Quarterly* 26:358; October 1955.
85. WHEAT, HARRY GROVE. *Foundations of School Learning.* New York: Alfred A. Knopf, 1955. 391 p.
86. WIEBE, VERNON R. "A Study of Tests of Kinesthesis." *Research Quarterly* 25:222; May 1954.
87. WITTE, FAE. *A Factor Analysis of Measures of Kinesthesis.* Doctor's thesis. Bloomington: Indiana University, 1953. (Microcard PH 20.)
88. WOODWARD, ROBERT S. *Experimental Psychology.* New York: Henry Holt & Co., 1949.
89. WORCHEL, PHILIP. "The Role of the Vestibular Organs in Space Orientation." *Journal of Experimental Psychology* 44:4; 1952.
90. WORCHELL, PHILIP. "The Vestibular Organs in Space Orientation." *Perceptual and Motor Skills* 35:164; 1955.
91. YOUNG, OLIVE G. "A Study of Kinesthesis in Relation to Selected Movements." *Research Quarterly* 16:277; December 1945.
92. YOUNG, OLIVE G. "Rate of Learning in Relation to Spacing of Practice Periods in Archery and Badminton." *Research Quarterly* 25:231; May 1954.
93. YUKIC, ELEANOR C. "Group Movement and Growth in a Physical Education Class." *Research Quarterly* 26:222; May 1955.

CHAPTER 12

Research and the Curriculum

LOUIS E. ALLEY
RALPH H. JOHNSON
WALLACE ANN WESLEY

THE CURRICULUM BUILDER MUST DETERMINE THE VALUES THAT are considered by society to be desirable and must attempt to define specific, fractional objectives that, when realized, will result in the attainment of those values. He must select the activities, the experiences, and the instructional materials through which the objectives may be reached. He must organize and arrange these activities, experiences, and instructional materials in such a manner that the objectives may be realized efficiently. He must evalua.. in terms of the declared objectives the effectiveness of the curriculum he proposes.

The problems faced by the curriculum builder are broader and less clearly defined than are the problems attacked by the worker engaged in "pure" research. The worker in pure research is concerned with testing specific hypotheses under rigidly controlled conditions. The researcher in curriculum works within a general area concerned with the nature and needs of contemporary society, the nature and needs of children and youth, and the nature of the learning process. To accomplish his task, the curriculum builder must employ a variety of techniques which directly or indirectly utilize most of the research methods described in this volume. He must philosophize concerning the nature of the "good life." He must gather historical and current statistics from a wide variety of sources, tabulate them, analyze them, and interpret them in terms of their implications for education. He must glean from the research in related fields of knowledge the findings that are appropriate to his purposes. He must formulate hypotheses, measure, experiment, measure again, and interpret the differences in his measurements. He must attempt to link effect with cause.

Krug (102:254) states that research as related to curriculum development means "the gathering and interpreting of evidence, from any source whatsoever, primary or secondary, that is useful and necessary in the solution of educational problems."

The diversity of the problems in which the curriculum builder is interested prohibits any but a cursory treatment in this chapter. An attempt has been made to acquaint the reader with the nature and scope of the problems faced by the curriculum researcher and, by classifying and documenting studies that are readily accessible, to guide the reader in gaining insight into the manner in which some of these problems have been attacked. The studies documented illustrate attempts to solve such problems. No effort has been made to describe and evaluate the research techniques employed.

Research and the Curriculum in Physical Education

LOUIS E. ALLEY

When one examines the research related to the development of the curriculum in physical education, sharply defined shifts of emphasis often cannot be chronologically identified. In general, emphases in such research have paralleled and lagged behind developments in the theory of general education. Once established, an emphasis continues to exert an influence upon subsequent research. To point out one of many examples, the attention drawn by educators to the importance of "interest" to the learning process (for an example, see 46) exerted a continuing influence upon research in physical education to the extent that some 30 studies in which interests in physical education were investigated appear in the *Research Quarterly*.

HISTORICAL TRENDS

It is possible, however, to describe in general terms the shifts in emphasis, and the developments in respect to research procedures employed, that appear to have played major roles in the evolution of the curriculum in physical education. An indication of the chronological sequence of the trends discussed may be found in the dates of the publications cited.

Articles in which programs of physical education as offered in specific schools are described appear in early issues of the *Research Quarterly* (28, 31, 143). Such articles probably motivated some teachers to examine their own programs critically and to modify them in instances in which unfavorable comparisons indicated a need for such modification.

Surveys to determine current practices in physical education were among the early attempts at research related to the curriculum. In 1933 an account was published of a comprehensive survey (19) of selected schools in which the curriculums in physical education were considered to be outstanding. Accounts of similar surveys (for an example, see 134) have appeared regularly in the literature up to the present date as researchers in curriculum have attempted to keep informed concerning current changes in the field.

Attempts to utilize the weight of expert opinion in shaping the curriculum have played a major role in research related to the curriculum. The Committee on Curriculum Research of the College Physical Education Association in 1930 published the first of a series of reports on a continuing study designed to develop a comprehensive graded program of physical education extending from the first grade through college (for summary of reports, see 108). The opinions of college and university experts in physical education, of state and city directors of physical education, and of physical education instructors were collected and analyzed and used as the basis for the proposed curriculum.

The development of tests of physical qualities (Chapter 8), improved statistical techniques (Chapter 7), and the utilization of research methods employed in physiology and in psychology (Chapter 11) enabled the curriculum worker to determine the effects of specific activities upon participants (11, 12, 51, 70, 71, 77, 92, 141, 176, 186, 188), the effects of programs of activities upon participants (105, 177, 181), and to compare the effectiveness of different types of programs (154, 185, 187).

Studies of the physical growth of children (118) and of the acquisition of motor skills as functions of maturation (47, 67, 114) were followed by studies of the roles played by growth and maturation in the performance of activities common to physical education programs (52, 53, 54, 55, 94, 103, 137, 164).

The advent of World War II not only exposed an alarming lack of fitness on the part of the male youth of the United States, but also created the need for programs that developed physical fitness to a high degree. In short order the physical education curriculums in high schools and particularly in colleges (151) were adapted to meet the needs of young men bound for war. As a result, the literature during the war years is liberally sprinkled with research studies to determine the effects of physical fitness programs upon the participants (16, 43, 74, 98, 131).

When instruments for evaluating emotional reactions, personality traits, and social adjustments were made available, researchers in physical education attacked the problem of attempting to determine the effects of participation in specific activities and in programs of physical education upon emotions, personality, and social adjustment (12, 85, 86, 87, 113, 155, 165, 167).

Under the direction of Bookwalter and others, a series of studies (133) was initiated to evaluate, by means of the LaPorte Health and Physical Education Score Card No. II, the physical education curriculums offered in each of the states. In addition to presenting an over-all picture of physical education in high schools throughout the United States, these studies afford opportunities for comparing the programs offered in the different states; for determining the relationships between the status of physical education in the high schools and such factors as school enrollment, size of community, accreditation, geographic area, type of school district, and consolidation; and for comparing, in terms of pupil achievement, the results obtained in high schools that offer good physical education programs with the results obtained in high schools that offer poor programs.

A study by Kraus and Hirschland (100), which disclosed that the "muscular fitness" of American youth compared unfavorably with that of European youth, unleashed a chain of events that resulted in the formation of President Eisenhower's Council on Youth Fitness (149) in July 1956. Many state committees and councils on fitness have been formed and the impact of nation-wide attention on fitness is being reflected in school programs of physical education.

A major share of the published research that relates to the curriculum in physical education is concerned with physical education at the college and university level. Very little attention has been given to research related to the physical education curriculum for the elementary and secondary schools. The growing emphasis upon action research (Chapter 13), in which problems related to the curriculum are studied in practical situations, is a promising development that may result in improved physical education curriculums in elementary and secondary schools.

RESEARCH IN CURRICULUM DEVELOPMENT

Biological Nature and Needs of Children and Youth. Research related to the biological nature and needs of children and youth provides basic information with which the curriculum builder may formulate guiding principles for establishing curriculums in physical education (for examples of such principles see 39: 145-146, 165-166, 186, 208-209, 237, 258-259). Research that is directly concerned with the effects of activities upon participants is also invaluable to the curriculum worker. Examples of both types of research are documented below.

Physical Growth and Motor Development. Because the activities that comprise the curriculum should be selected and assigned to grade levels in accordance with the physical growth and motor development of the children and youth for whom the curriculum is intended, basic studies of growth and development such as those reported by Baldwin (6), McGraw (114), and Meredith (118) are of particular interest to the curriculum researcher.

Espenschade (52), in a comprehensive study that was a part of the Adolescent Growth Study of the Institute of Child Welfare of the University of California, investigated the relationship between a selected group of motor functions and anatomical and physiological development. She has also reported studies that deal with the role of physiological maturity in physical activities (54), changes in co-ordination with changes in age (53), and the effects

1

of rapid growth in adolescent boys upon dynamic balance (55). Dimock (47), in a study of adolescent boys, investigated the relationship of pubescence and growth in height and weight, the influence of age on physical growth, and the relationship of strength and motor ability to pubescence. These studies provide basic material for the curriculum builder.

Keeler (94), in a study of boys in grades 5 through 12, investigated the relationship between maturation and performance on the Johnson Skill Test. Nevers (137) studied the effect of the various cycles of maturity (prepubescence, pubescence, postpubescence) upon the ability of junior high school boys to perform selected motor skills. Jones (88) conducted a longitudinal study to determine differences in strength for premenarcheal and postmenarcheal girls of the same chronological ages. Seils (164) assessed carpal X rays to determine the maturity of girls and boys six to eight years old and attempted to determine relationships between maturity, physical growth, and proficiency in performing selected gross motor skills.

Sex Differences. Moore (127) summarized the findings of research workers in regard to sex differences that affect performance in physical activities. The findings are classified according to anatomical differences, physiological differences, and psychological differences.

Physiological Effects of Exercise and of Participation in Specific Activities and in Programs. Among the unique contributions made by physical education toward the accomplishment of the objectives of general education is the development of physical fitness. A knowledge of the physiological effects of exercise and of participation in specific activities and in programs of activities is essential to sound curriculum construction.

The curriculum builder will find such comprehensive summaries of the effects of exercise as that prepared by Steinhaus (171) to be invaluable as sources of information from which basic principles for curriculum planning may be derived. The publications of Hellebrandt (63, 64) also provide such information.

Studies concerning the effects of exercises of graded intensity include Shirley's study (166) of the response of the normal prepubescent heart to graded exercises; Hogdson's study (69) of respiratory and circulatory reactions to exercise; and Meyer's and

Tuttle's studies (120, 176) of the effects of graded exercise upon the leukocyte count.

Numerous studies have been conducted to determine the physical effects of participation in specific activities. Such studies present data concerning: (a) the effects of weight training upon such factors as speed of movement, co-ordination, and power (30, 117, 188, 186); (b) the effects of horizontal-ladder exercises upon the upper body strength of third-grade children (77); (c) the relationship between the frequency of play periods and the time devoted to play to such factors as physical fitness and motor ability (178); (d) the effects of participation in water polo upon blood pressure and pulse rate (51); (e) the effects of participation in interscholastic basketball upon the physical fitness of high school boys (141); (f) the effects of a season of training and competition in track and field athletics upon the hearts of high school boys (177); (g) the effects of participation in eight selected physical activities upon the physical fitness and motor ability of college freshmen (105); (h) the effects of participation in two-court and three-court basketball upon the respiratory rates, metabolism, pulse rates, systolic and diastolic pressures of college women (70, 71); (i) the effects of modern dance, folk dance, basketball, and swimming upon the development, the agility, the strength, the flexibility, the power, and the general motor ability of college women (11); (j) the effects of strenuous exercise programs upon the physical efficiency of college women (181); and (k) the effects of participation in marching, free exercises, dance, and a tag game (poison) upon the pulse rate of college women (150).

Cureton (40) has assembled in mimeographed form summaries of 227 unpublished theses in which the effects of participation in physical education and athletics upon college men are reported.

Psychological Nature and Needs of Children and Youth. The curriculum builder should plan a curriculum in physical education in accordance with the psychological nature and needs of the children and youth for whom the curriculum is intended. Basic research in interests, attitudes, emotions, and social adjustment are of particular concern to the researcher in curriculum. This concern is exemplified by a summary of studies that relate to adolescents, together with the implications of these studies for the curriculum in physical education (29).

Interests. In most of the studies of the interests of children and youth in activities, researchers have utilized the questionnaire and/or interview method for obtaining data. Attempts have been made to link interests and preferences of students to past experiences (8, 9, 23, 26, 174); to physical qualities such as stature, strength, and motor ability (5, 33, 48); to administrative policies in physical education (26, 45, 90); and to the training of high school teachers of physical education (26).

Cowell (37) utilized the diary analysis technique to determine the activities that junior high school boys find interesting and worthwhile. Alden (1) investigated the factors in the required physical education program that are least desirable to college women.

A major share of the studies of interests in, and preferences for, activities are concerned with students of college age. Studies related to interests and preferences of college women are more abundant than such studies related to college men. Very few such studies related to elementary school children appear in the literature; among those few studies are those by Lehman and Witty (112), by Blanchard (13), and by the Committee for Elementary School Boys and Girls of the New York State Physical Education Standards Project (139).

Attitudes. Lapp (110) analyzed the responses of high school students to a questionnaire to determine the values the students expected to derive from physical education. Cowell, Daniels, and Kenney (38) studied, by means of a checklist, the values that university freshmen and directors of service programs sought in service programs and the values approved by college presidents. Moore (125) used Form A of the Bues-Remmer Scale, supplemented by interviews, to evaluate the attitudes of college women toward physical activity as a means of recreation.

Kappes (91) developed an inventory for determining attitudes of college women toward physical education and student services in the physical education department. Wear (182) developed an attitude inventory for evaluating attitudes of college men toward physical education as an activity course and later constructed equivalent forms of the inventory to provide instruments for measuring changes in attitude that result from intervening experiences (183). The Wear Attitude Inventory was subsequently used to determine the attitudes of college women toward physical education (10, 22).

Emotions, Personality, and Social Adjustment. To evaluate emotional responses of college athletes, Johnson utilized changes in heart rate, blood pressure, and blood sugar (85); psychogalvanic and word association techniques (86); and Buck's House-Tree-Person Test (87). Skubic (167) compared the emotional responses of boys participating in Little League Baseball and of boys participating in softball in physical education classes.

Blanchard (12) developed a "behavior frequency rating scale" for analyzing the character and personality traits that high school boys and girls exhibited in physical education classes. Seymour (165) used the Science Research Associates Junior Inventory, the Winnetka Scale for Rating School Behavior and Attitudes, and the Ohio Social Acceptance Scale for the Intermediate Grades to study behavior characteristics of participants and nonparticipants in Little League Baseball. Reid (155) studied the contributions that the freshman year in a liberal arts college for women made to personality as measured by the Minnesota Multiphasic Personality Inventory.

McCraw and Tolbert (113) studied the results obtained on sociometric tests and on tests of general athletic ability to determine the relationship of social status to athletic ability.

The Process of Learning. The curriculum should be organized and arranged in temporal sequence in a manner conducive to efficient learning and to meeting the needs of all the students. Consequently, researchers in curriculum are interested in basic research in the psychology of motor learning from which implications for physical education may be drawn, in studies of individual differences in pupils, and in studies related to the grade placement of activities. Published research of this nature appears to be rather meager in the area of physical education.

Psychological Theory Applied to Physical Education. Schwendener (159) discusses, with particular reference to time allotments for instruction and to methods of teaching, the application of educational theory to physical education. The discussion by Cornwell (36) of the psychology of motor learning has some implications for the curriculum builder.

Individual Differences. The studies by Rarick and McKee (152), in which differences and similarities in children of high and low levels of motor achievement were studied, and by Cowell (37), in which consideration is given to the "fringers," are examples of types of research in physical education that are needed to provide basic information on individual differences for the curriculum builder.

Grade Placement of Activities. Investigations designed to determine the grade placement of activities in terms of the efficiency with which the activities are learned appear infrequently in research in physical education. In studies of the effects of instruction in throwing upon the throwing ability of young children (49, 121), no improvement in accuracy as a result of such instruction was reported, but improvement in the distance of the throw was reported.

The most common procedure for attempting to determine the grade placement of activities has been to utilize the weight of opinions of experienced teachers (50, 153) and experts in physical education (109).

Nature and Needs of Contemporary Society. Research studies in physical education that are related to the nature and needs of contemporary society and that have implications for the curriculum include surveys of the status of current curriculums in physical education; surveys of the duties of teachers and coaches (e.g., job analyses, combinations of subjects taught); studies of problems of beginning teachers; analyses of certification requirements for teachers; analyses of the qualities regarded important by those who employ teachers; and studies that reveal needs or deficiencies for which physical education may be held accountable.

Current Status of Curriculums. Numerous questionnaire surveys of current practices in physical education may be found in the literature. These studies include surveys of elementary school curriculums (60); junior high school curriculums (81); high school curriculums (19, 35, 78, 83, 99); service programs in colleges and universities (7, 44, 151); coeducational programs of physical education (34, 101); intramural programs in colleges (59, 111); professional curriculums for undergraduates (3, 82, 84); and graduate programs (89, 140).

As a step in developing recommendations for professional curriculums for undergraduates, Peik and Fitzgerald (145) and Neilson (135, 136) analyzed course offerings listed in college catalogues.

Duties of Teachers and Coaches. Important data to be considered in formulating the professional program in physical education may be obtained from studies concerning the duties performed by teachers. In such studies the duties of teachers are usually classified in terms of the enrollments of schools in which the duties are required (79); appropriateness of including in the teacher-training program preparation for such duties (68); or frequency, difficulty, and importance of such duties (76, 116, 163).

Studies concerning the subject combinations taught by teachers of physical education provide information that is useful in determining suitable minors for physical education majors. Such studies have been reported by Rugen (156), Street (173), Horton (73), and Moore (126).

Problems of Beginning Teachers. Information gained from studies of problems encountered by beginning teachers may be used to determine points for emphasis in the teacher-training program (18, 93). Brown (25) compared the problems encountered by student teachers with the amount of attention given to the problems in textbooks on methods of teaching.

Analysis of Certification Requirements. Studies by Morehouse and associates (128, 129, 130), in which the requirements for teacher certification in physical education are presented, provide information necessary to formulating the professional program for undergraduates.

Qualities Regarded as Important by Employers. In prescribing professional programs for undergraduates, attention should be given to developing in students the qualities desired by employers. Reports designed to provide such information include the viewpoint of a state director of physical education (175), the viewpoint of a school administrator (20), and a summary of the opinions held by administrators and principals in large towns and cities (61).

Needs or Deficiencies. Moffett (123) conducted a questionnaire study designed to provide information useful in formulating a graduate program to meet the needs of the teachers most likely to attend summer sessions.

The report of Kraus and Hirschland (100) to the effect that American children as compared with European children are deficient in muscular fitness is a classical example of data useful in the "shortage approach" to curriculum construction. Wendler (184) used the shortage approach in recommending revisions in service programs and in physical efficiency standards for college students.

Research of a Philosophical Nature. The disagreement among scholars concerning the place of philosophical studies (Chapter 15) in scientific research is reflected in the literature by a dearth of studies of a philosophical nature—a dearth that is unfortunate for the curriculum builder. Philosophical studies point out the goals toward which the curriculum should lead; facts supplied by scientific research serve only as guides for reaching those goals.

The curriculum should be determined by the values regarded as desirable by society. The selection of such values is a matter of choice. However, in choosing values and in developing a curriculum designed to attain those values, the researcher in curriculum should systematically assemble, study, interpret, and apply all pertinent facts—a procedure that necessarily involves philosophical considerations.

Determination of Objectives. The Committee on Curriculum Research of the College Physical Education Association, in attempting to determine the objectives for physical education, initiated a study (107) in which a variation of the "pooled thinking" method described by Cureton (41) was employed. The committee collected the objectives listed in books, state courses of study, municipal courses of study, reports of national professional committees, and professional journals; determined the frequency with which the objectives were listed; classified the objectives under four headings; and established criteria for selecting worthy objectives.

In establishing for each grade level the educational, emotional, physiological, and social objectives for each activity in the curriculum for elementary school and junior high school students, the Research Committee of the Newark Physical Education Association (138) utilized a procedure similar to that described above.

Establishment of Principles or Standards. Esslinger (56), in establishing principles for selecting activities in physical education, utilized facts drawn from anatomy, physiology, psychology, and education that are related to the growth, development, capacities, and interests of children, and facts derived from a study of contemporary society and social trends that are related to the needs of children and adults.

Meshizuka (119) developed a program of professional training in physical education for colleges and universities in Japan by utilizing guiding principles based upon a critical review of the literature pertaining to professional programs; a consideration of the current program of professional training in physical education in Japan; geographical, socioeconomic, biological, and pedagogical factors peculiar to Japan that condition the nature of professional training; and a survey of programs of professional training in physical education in 13 institutions in the United States and in 15 countries outside the United States.

Interpretation in Terms of Philosophies. In a study unique in physical education literature, Clark (32) attempted an interpretation of a college program in terms of realism, pragmatism, and idealism in which she described the basic tenets of each philosophy and examined the parts of the physical education program for evidence of the influence of each philosophy.

EVALUATION OF THE CURRICULUM

Having examined and selected values, determined objectives, and prescribed activities, the researcher in curriculum then should evaluate the effectiveness of the curriculum he has prescribed. His evaluation may be made in terms of the values he had selected as desirable or the objectives he had set out to accomplish through the curriculum he prescribed. He may evaluate, in the direction of the objectives or the values, the progress made by the pupils, the achievement level attained by the pupils, or both. In making his evaluation, he may utilize objective measurements and rigid statistical procedures, subjective methods, or various combinations of both. Having evaluated the curriculum, he may compare its effectiveness with that of other curriculums similarly evaluated.

Elementary School and High School Programs. Reports of evaluations of elementary school and high school programs in physical education are rather limited in both number and scope. Kelly (95) utilized the LaPorte Score Card No. I, personal interviews, and a supplementary questionnaire to evaluate the elementary school program offered in the public schools of Lafayette, Indiana.

Rath (154) compared the effectiveness of three types of physical education curriculums for ninth-grade boys by comparing the gains in strength made by boys in the various programs. Bookwalter (17) describes the attempts made by Bonsett in Indiana high schools and by DeVoll in Wisconsin high schools to determine the relationship between the standards governing the administration of physical education programs (as measured by the Health and Physical Education Score Card No. II) and the achievement of the objectives of physical education (as measured by selected tests of physical fitness, sports skills, sports knowledge, and attitudes).

Several self-appraisal checklists by means of which persons or committees may evaluate programs of physical education in secondary schools have been developed, usually by state or national committees. Daniels, at the direction of the Ohio Association for Health, Physical Education, and Recreation, developed such a checklist (42) for use in Ohio secondary schools. The items included in the checklist reflect the thinking of selected specialists who were familiar with physical education in the secondary schools of Ohio. The American Council on Education, through the Committee on Cooperative Study of Secondary School Standards, developed criteria that may be used for evaluating all phases of the secondary school curriculum, including physical education for boys and physical education for girls (4).

Service Programs in Colleges and Universities. During World War II, the development of physical fitness was emphasized in most service programs for men in colleges and universities. Evaluations of such programs were usually made in terms of gains in physical fitness or in terms of levels of physical fitness achieved by participants (16, 43, 74, 98, 131).

Phillips (146), in evaluating the service programs in liberal arts and teachers colleges of New York, determined from the statements of authorities and from a survey of authoritative literature the needs of college students that should be fulfilled through physical education, and the basic principles that should govern the operation of the physical education program. Program standards based on these needs and basic principles were submitted to a panel of 12 recognized authorities in the field of physical education, and optimal and essential standards were established. Data concerning the programs offered in colleges and universities were obtained by means of questionnaires, the reliability of which were determined by visits to selected institutions.

Wilbur (185) compared the effectiveness, in terms of changes in Physical Fitness Indices, of a gymnasium-type program and a sports-type program.

Kenney (97) evaluated the effectiveness of the required physical education program for men at the University of Illinois in terms of the leisure habits of the University of Illinois graduates.

Adapted Programs. Landers (104) developed a score card with which to evaluate physical education programs for physically handicapped children in public schools. The score card is based upon the needs of such children as determined by a survey of the literature and from the results of a questionnaire survey of the opinions of experts in the areas of medicine, orthopedics, physical therapy, correctives, and adapted physical education.

Broer (21) determined the effectiveness of a basic skills curriculum for women of low motor ability by comparing the levels of achievement in skills and knowledge, and the changes in motor ability and attitude of two matched groups of students. The first group, before entering the service program had participated in a basic skills curriculum; the second group entered directly into the regular service program.

Professional Curriculums. Research in which the evaluation of professional curriculums is a primary objective includes: (a) the development of criteria based on an analysis of certification requirements in each of the states, an analysis of the professional requirements in selected schools that offered curriculums for majors in physical education, and the opinions of experts in the field of physical education (15); (b) evaluations, by teachers, of the adequacy of the training they received at the institutions they attended (14, 80); and (c) a comparison of the changes recommended by educators and the changes that actually occurred in state teachers college curriculums (157).

Campbell (27) evaluated one aspect of the professional curriculum by comparing the scores made by physical education majors on the American Council of Education Contemporary Affairs Test for College Students with national norms for such scores and with scores made by majors in areas other than physical education.

Standards for evaluating institutions that offer professional curriculums in physical education have been formulated by national committees. As the basis for developing standards for a four-year undergraduate curriculum and a three-year graduate curriculum, the National Committee on Standards (132) determined the basic characteristics of secondary school curriculums of physical education and analyzed the types of duties required of physical education teachers. This committee also formed a National Rating Committee.

In 1952 the National Continuing Committee for the Improvement of Professional Preparation in Health Education, Physical Education, and Recreation, in co-operation with the Committee on Studies and Standards of the American Association of Colleges for Teacher Education (currently called the National Council for Accreditation of Teacher Education) developed evaluation schedules in health, physical education, and recreation that may be used as self-evaluating instruments or by visitation teams in evaluating professional curriculums. *Evaluation Standards and Guide,* a booklet published by the American Association for Health, Physical Education, and Recreation in 1959, is a revision of these evaluative criteria for college and university programs.

DEVELOPMENT OF INSTRUCTIONAL MATERIALS

Research ventures in the development of instructional materials have been largely limited to the construction of knowledge tests in physical education activities; the classification, evaluation, and development of films suitable as instructional aids in physical education; and the co-operative development of a textbook in physical education for high school students.

Knowledge Tests. Few reports of studies designed to produce knowledge tests in physical education for elementary school and high school students appear in the research literature. (See Chapter 8 for a comprehensive discussion of knowledge tests.) Heath and Rodgers (62) developed a knowledge and skills test in soccer for fifth- and sixth-grade boys. Schwartz (158) developed T-scales based upon the results of scores made by high school girls on knowledge tests in girls basketball. Stradtman and Cureton (172) prepared a physical fitness knowledge test for secondary school boys and girls. It is designed to measure a knowledge of desirable practices in developing and maintaining fitness.

Knowledge tests in physical education for college students have received a considerable amount of attention from the research worker. Reports of such tests provide measuring instruments designed to test the student's knowledge of a single activity (24, 57, 66, 147, 160, 161, 162, 179, 180) and batteries of tests covering a number of activities (65, 168, 169, 170). For a limited number of activities, knowledge tests designed for physical education majors are available (58, 96, 106, 122).

Films. Hughes and Stimson (75) developed a classified list of films related to health and physical education that are available to teachers. Payne (144) constructed a rating scale for the evaluation of such films and developed a catalogue in which selected films suitable for use in classes in physical education for girls are listed.

Homewood (72) produced a sound film for use in teaching skills commonly presented in beginning classes in girls basketball. Owens (142) produced a film designed to provide a training experience for in-service and prospective elementary school classroom teachers. Porter (148) produced a similar film for special teachers of physical education in the primary grades.

Textbook for High Schools. The co-operative development and publication of a textbook in physical education for high school students (124) marks the first attempt in physical education to provide comprehensive, standardized reference materials for such students. The impact of the textbook upon secondary school curriculums has not been determined (for a discussion of the possible impact upon the curriculum, see 2).

NEEDED RESEARCH IN CURRICULUM

The researcher who is interested in curriculum construction will not lack problems to solve. The thoughtful student, noting the research efforts that have been made in attempts to solve problems related to the curriculum, will recognize that many problems remain unsolved. The total picture is far from a complete one. Evidence that will justify many parts of the curriculum is either inconclusive or missing. Considerable controversy exists concerning the manner in which the known parts should be fitted into a unified pattern.

If the evidence to support the curriculum was complete and clear in every detail, the researcher in curriculum would still find plenty of work to be done. The curriculum cannot long remain static because the society for which the curriculum prepares children and youth is constantly changing. The curriculum should be continually evaluated in terms of meeting the needs of the children and youth it is intended to serve, and changes indicated by the results of the evaluations should be promptly made.

Below are listed some suggestions for research studies which, in view of the research completed, would appear to be both fruitful and needed:

1. Studies in which the results of research already completed are summarized, evaluated, and interpreted in terms of practical implications for the physical education curriculum. Particular attention should be given to research completed in the areas of child development, psychology, and sociology, as well as to research in physical education.
2. Studies of a philosophical nature in which attempts are made to determine the values to be sought through the physical education curriculum, together with suggestions of means by which those values might be attained
3. Longitudinal studies of the development of strength, endurance, and fundamental motor skills as measured by tests commonly used in physical education, with particular attention given to individual developmental patterns
4. Studies related to such problems associated with the organization of the curriculum as (a) grade placement of activities in terms of difficulty of activities and pupil readiness, (b) vertical sequence with respect to each activity and with respect to all the activities included in the program, (c) distribution of time allotments, (d) provisions for individual differences, and (e) activities that should be required and activities that should be elective
5. Studies in which the effectiveness of the use of curriculum materials (textbooks, films, charts) is evaluated
6. Studies in which the curriculum is evaluated in terms of permanent effects on participants
7. Studies related to physical education in the elementary school
8. Studies in which the roles of students and lay people in curriculum development are investigated.

BIBLIOGRAPHY ON THE CURRICULUM IN PHYSICAL EDUCATION

1. ALDEN, MABEL ANNA. "The Factors in the Required Physical Education Program Least Attractive to the College Girl." *Research Quarterly* 3:97-107; December 1932.
2. ALLEY, LOUIS E. "Welcome Stranger—New High School Physical Education Textbook." *Journal of Health-Physical Education-Recreation* 26:6:8-9; September 1955.

3. ALLMAN, DELMAR I. "Undergraduate Requirements in Anatomy and Physiology for Physical Education Majors in Eighty Three Colleges and Universities." *Research Quarterly* 27:117-19; March 1956.
4. AMERICAN COUNCIL ON EDUCATION. Cooperative Study of Secondary School Standards. *Evaluative Criteria.* Washington, D. C.: the Council, 1950. Section D-13: Physical Education for Boys; Section D-14: Physical Education for Girls.
5. ANDERSON, THERESA W. "The Attitudes of High School Girls Toward Physical Activities." *Research Quarterly* 5:48-61; December 1934.
6. BALDWIN, BIRD T. *The Physical Growth of Children from Birth to Maturity.* University of Iowa Studies in Child Welfare. Vol. 1, No. 1. Iowa City: State University of Iowa, 1921.
7. BARR, MARGARET C. "Résumé of the Physical Education Activity Program in Colleges for Women." *Research Quarterly* 4:117-31; October 1933.
8. BEISE, DOROTHY. "A Comparative Analysis of the Physical Education Background, Interests, and Desires of College Students as an Evaluation Procedure." *Research Quarterly* 11:120-34; December 1940.
9. BEISE, DOROTHY. "Physical Education Background of College Students as a Factor in Determining the Content of the Required Physical Education Program." *Research Quarterly* 6:39-52; May 1935.
10. BELL, MARGARET; WALTERS C. ETTA; and STAFF. "Attitudes of Women at the University of Michigan Toward Physical Education." *Research Quarterly* 24: 379-91; December 1953.
11. BENNETT, COLEEN L. "Relative Contributions of Modern Dance, Folk Dance, Basketball, and Swimming to Motor Abilities of College Women." *Research Quarterly* 27:253-62; October 1956.
12. BLANCHARD, B. EVERARD. "A Comparative Analysis of Secondary-School Boys' and Girls' Character and Personality Traits in Physical Education Classes." *Research Quarterly* 17:33-39; March 1946.
13. BLANCHARD, V. S. "Analysis of the Likes and Dislikes of Boys and Girls in Health Education Activities." *Research Quarterly* 4:238-45; March 1933.
14. BLANCHARD, V. S. "Guiding Principles in Teacher Training from a Public School Point of View." *Research Quarterly* 3:81-92; May 1932.
15. BLESH, T. ERWIN. "Evaluation Criteria in Physical Education." *Research Quarterly* 17:114-26; May 1946.
16. BOOKWALTER, KARL W. "A Critical Analysis of Achievements in the Physical Fitness Program for Men at Indiana University." *Research Quarterly* 14: 184-93; May 1943.
17. BOOKWALTER, KARL W. "Physical Education Educates!" *Physical Educator* 14: 43-44; May 1957.
18. BRAINARD, A. F. "A Survey of the Problems Confronting Men Student Teachers in the Field of Physical Education." *Research Quarterly* 13:332-40; October 1942.
19. BRAMMELL, P. ROY. *Health Work and Physical Education.* U. S. Department of Health, Education, and Welfare, Office of Education. National Survey of Secondary Education, Monograph No. 28. Washington, D. C.: Superintendent of Documents, Government Printing Office, 1933.
20. BRISTOW, WILLIAM H. "The Training of Teachers from the Viewpoint of a School Administrator." *Research Quarterly* 3:99-106; May 1932.
21. BROER, MARION R. "Evaluation of a Basic Skills Curriculum for Women Students of Low Motor Ability at the University of Washington." *Research Quarterly* 26:15-27; March 1955.
22. BROER, MARION R.; FOX, KATHERINE S.; and WAY, EUNICE. "Attitude of University of Washington Women Students Toward Physical Education Activity." *Research Quarterly* 26:379-84; December 1955.
23. BROER, MARION R., and HOLLAND, DOLLY A. J. "Physical Education Interests and Needs of University of Washington Women in Service Classes." *Research Quarterly* 25:387-97; December 1954.

24. BROER, MARION R., and MILLER, DONNA MAE. "Achievement Tests for Beginning and Intermediate Tennis." *Research Quarterly* 21:303-13; October 1950.
25. BROWN, MARGARET C. "Learning Problems in Student Teaching." *Research Quarterly* 11:25-32; December 1940.
26. BULLOCK, MARGUERITE, and ALDEN, FLORENCE D. "Some of the Factors Determining the Attitude of Freshmen Women at the University of Oregon Toward Required Physical Education." *Research Quarterly* 4:60-70; December 1933.
27. CAMPBELL, LAURIE E. "The Prospective Teacher and His Knowledge of Contemporary Life Affairs." *Research Quarterly* 11:126-35; October 1940.
28. CARTER, JAMES J. "The Lynbrook Physical Education Program." *Research Quarterly Supplement* 6:84-94; March 1935.
29. CASSIDY, ROSALIND. "Recent Studies of American Adolescents as Guides for the Reorganization of the Secondary School Physical Education Program." *Research Quarterly* 9:38-42; December 1938.
30. CHUI, EDWARD. "The Effect of Systematic Weight Training on Athletic Power." *Research Quarterly* 22:188-94; March 1951.
31. CLARK, LEONARD. "Melrose High School Experiments." *Research Quarterly Supplement* 6:111-15; March 1935.
32. CLARK, MARGARET C. "A Philosophical Interpretation of a Program of Physical Education in a State Teachers College." *Research Quarterly* 15:325-29; December 1944.
33. CLEVETT, MELVIN A. "The Interests and Participation of Boys in Voluntary Physical Education Activities." *Research Quarterly* 5:27-42; October 1944.
34. COBB, LOUISE S. "The Co-Educational Physical Education Class Program at the College Level." *Research Quarterly* 10:20-32; December 1939.
35. COLLINS, LAURENTINE B., CHAIRMAN. "Curriculum Study for Public Schools Section of the American Physical Education Association." *Research Quarterly* 5:111-26; December 1934.
36. CORNWELL, OLIVER K. "The Psychology of Motor Skills." *Research Quarterly* 4:126-30; March 1933.
37. COWELL, CHARLES C. "Diary Analysis: A Suggested Technique for the Study of Children's Activities and Interests." *Research Quarterly* 8:158-72; May 1937.
38. COWELL, CHARLES C.; DANIELS, ARTHUR S.; and KENNEY, HAROLD E. "Purposes in Physical Education as Evaluated by Participants, Physical Education Supervisors, and Educational Administrators." *Research Quarterly* 22:287-97; October 1951.
39. COWELL, CHARLES C., and HAZELTON, HELEN W. *Curriculum Designs in Physical Education.* New York: Prentice-Hall, 1955.
40. CURETON, THOMAS KIRK, EDITOR. "Effects of Physical Education and Athletics upon College Men." Urbana: University of Illinois, 1955. (Mimeographed)
41. CURETON, THOMAS KIRK. "The Philosophical or Group Thinking Method of Research." *Research Quarterly* 11:75-83; October 1940.
42. DANIELS, ARTHUR S., CHAIRMAN. *Evaluative Criteria for Physical Education.* Ohio Association for Health, Physical Education, and Recreation. Columbus: Division of Health, Physical Education, and Recreation, Ohio State Department of Education, 1957.
43. DAUBERT, R. B. "A Physical Conditioning Program as Conducted at Michigan State College." *Research Quarterly* 14:176-83; May 1943.
44. DAVIS, DOROTHY. "Nation-Wide Survey of Core Curricular Content in Physical Education for College Women." *Research Quarterly* 25:144-49; May 1954.
45. DAVIS, ELWOOD CRAIG. "A Study of the Interests of the Pennsylvania State College Freshmen in Certain Formal and Natural Physical Activities." *Research Quarterly* 4:49-59; December 1933.
46. DEWEY, JOHN. *Interest and Effort in Education.* Boston: Houghton-Mifflin Co. 1913.
47. DIMOCK, HEDLEY S. "A Research in Adolescence: I. Pubescence and Physical Growth." *Child Development* 6:176-95; September 1935.

48. DRIFTMIER, ERNA. "Individual Differences in Interests and Physical Traits as Related to High School Girls in Physical Education." *Research Quarterly* 4: 198-220; March 1933.
49. DUSENBERRY, LOIS. "A Study of the Effects of Training in Ball Throwing by Children Ages Three to Seven." *Research Quarterly* 23:9-14; March 1952.
50. EKDAHL, GEORGE C. "A Study of the Most Popular Games of Low Organization." *Research Quarterly* 10:120-23; March 1939.
51. ELHARDT, W. P., and ORTH, O. S. "The Effect of Water Polo on Blood Pressure and Pulse Rate." *Research Quarterly* 7:99-101; May 1956.
52. ESPENSCHADE, ANNA. *Motor Performance in Adolescence, Including the Study of Relationships with Measures of Physical Growth and Maturity.* Monographs of the Society for Research in Child Development. Vol. 5, No. 1. Washington, D. C.: National Research Council, 1940.
53. ESPENSCHADE, ANNA. "Development of Motor Co-ordination in Boys and Girls." *Research Quarterly* 18:30-44; March 1947.
54. ESPENSCHADE, ANNA. "Physiological Maturity as a Factor in the Qualifications of Boys for Physical Activity." *Research Quarterly* 15:113-17; May 1944.
55. ESPENSCHADE, ANNA; DABLE, ROBERT R.; and SCHOENDUBE, ROBERT. "Dynamic Balance in Adolescent Boys." *Research Quarterly* 24:270-75; October 1953.
56. ESSLINGER, ARTHUR ALBERT. "A Philosophical Study of Principles for Selecting Activities in Physical Education." Doctor's thesis. Iowa City: State University of Iowa, 1938.
57. FOX, KATHARINE. "Beginning Badminton Written Examination." *Research Quarterly* 14:135-46; May 1953.
58. FRENCH, ESTHER. "The Construction of Knowledge Tests in Selected Professional Courses in Physical Education." *Research Quarterly* 14:406-24; December 1943.
59. GALLIGAN, G. E. "The Status of Intramural Athletics for Men in State Teachers Colleges." *Research Quarterly* 9:58-62; May 1938.
60. GEORGIADY, ALEXANDER, and SAVAGE, RUSSELL. "Status of Physical Education in Elementary Schools." *Research Quarterly* 11:40-46; May 1940.
61. GRAYBEAL, ELIZABETH. "A Consideration of Qualities Used by Administrators in Judging Effective Teachers of Physical Education in Minnesota." *Research Quarterly* 12:741-44; December 1941.
62. HEATH, MARJORIE L., and RODGERS, ELIZABETH G. "The Study of the Use of Knowledge and Skills Tests in Soccer." *Research Quarterly* 3:33-53; December 1932.
63. HELLEBRANDT, FRANCES A., and MEYER, MARGARET H. "Physiological Data Significant to Participation by Women in Physical Activities." *Research Quarterly* 10:10-23; March 1939.
64. HELLEBRANDT, F. A. "Physiology and the Physical Educators." *Research Quarterly* 11:12-29; October 1940.
65. HENNIS, GAIL M. "Construction of Knowledge Tests in Selected Physical Education Activities for College Women." *Research Quarterly* 27:301-309; October 1956.
66. HEWITT, JACK E. "Comprehensive Tennis Knowledge Test." *Research Quarterly* 8:74-84; October 1937.
67. HICKS, J. A. "The Acquisition of Motor Skills of Young Children." *Child Development* 1:90-105; June 1930.
68. HINDMAN, DARWIN A. "An Analysis of the Activities of Physical Education Teachers in High Schools." *Research Quarterly* 8:117-31; May 1937.
69. HODGSON, PAULINE. "Studies in the Physiology of Activity: I. On Certain Reactions of College Women to Measured Activity." *Research Quarterly* 7:3-25; March 1936.
70. HODGSON, PAULINE. "Studies in the Physiology of Activity: II. On Certain Reactions of College Women Following Participation in Two-Court Basketball." *Research Quarterly* 7:45-55; May 1936.

71. HODGSON, PAULINE. "Studies in the Physiology of Activity: III. On Certain Reactions of College Women Following Participation in Three-Court Basketball." *Research Quarterly* 10:53-60; October 1939.
72. HOMEWOOD, JEAN MARGARET. "A Film and Manual for Use in Developing Ability To Teach Sports Skills." Doctor's thesis. Iowa City: State University of Iowa, 1955.
73. HORTON, CLIFFORD E. "A Report on Teaching Combinations in Physical Education in the High Schools of Illinois for Men and Women." *Research Quarterly* 4:51-61; October 1933.
74. HUGHES, BRYAN O. "Test Results of the University of Michigan Physical Conditioning Program, June 15-September 26, 1942." *Research Quarterly* 13:493-511; December 1942.
75. HUGHES, WILLIAM L., and STIMSON, PAULINE E. "Motion Pictures in Health and Physical Education." *Research Quarterly* 9:104-53; March 1938.
76. HUMPHREY, JAMES H. "A Job Analysis of Selected Public School Physical Education Directors." *Research Quarterly* 24:56-66; March 1953.
77. HUTINGER, PAUL W. "Effect of Systematic Horizontal-Ladder Exercises Upon Upper Body Strength of Third-Grade Children." *Research Quarterly* 26: 159-62; May 1955.
78. IRWIN, LESLIE W., and REAVIS, WILLIAM C. "Practices Pertaining to Health and Physical Education in Secondary Schools." *Research Quarterly* 11:93-109; October 1940.
79. JACKSON, C. O. "Activities Engaged in by Teachers of Physical Education in High Schools in Illinois." *Research Quarterly* 13:242-51; May 1942.
80. JACKSON, C. O. "Activities Engaged in by Teachers of Physical Education in the High Schools of Illinois, Part II." *Research Quarterly* 14:60-81; March 1943.
81. JACKSON, C. O. "Physical Education in the Junior High Schools of Illinois." *Research Quarterly* 10:124-34; March 1939.
82. JACKSON C. O. "Recent Changes in Student Teaching Curricula and Major Problems in Teacher Training at 43 Institutions of Higher Learning." *Research Quarterly* 7:108-19; May 1936.
83. JACKSON, C .O. "The Status of Physical Education in the Accredited Secondary Schools of Illinois." *Research Quarterly* 9:47-59; March 1938.
84. JACKSON, C. O. "A Study of Current Practices in Student Teaching and Supervision." *Research Quarterly* 3:152-58; March 1932.
85. JOHNSON, WARREN R. "Emotion Revealed in Two Types of Athletic Contests." *Research Quarterly* 20:72-77; March 1949.
86. JOHNSON, WARREN R. "Psychogalvanic and Word Association Studies of Athletes." *Research Quarterly* 22:427-33; December 1951.
87. JOHNSON, WARREN R., and HUTTON, DANIEL C. "Effects of a Combative Sport as Measured by a Projective Test." *Research Quarterly* 26:49-53; March 1955.
88. JONES, HAROLD E. "The Sexual Maturing of Girls as Related to Growth in Strength." *Research Quarterly* 28:135-43; May 1957.
89. JONES, LLOYD M. "The Graduate Minor in Physical Education." *Research Quarterly* 19:18-21; March 1948.
90. KANE, ISABEL B., and HODGSON, PAULINE. "A Study of Factors Influencing Participation of College Women in an Elective Program of Physical Education." *Research Quarterly* 10:57-66; March 1939.
91. KAPPES, EVELINE E. "Inventory to Determine Attitudes of College Women Toward Physical Education and Student Services of the Physical Education Department." *Research Quarterly* 25:429-38; December 1954.
92. KARPOVICH, PETER V. "The Effect of Basketball, Wrestling, and Swimming Upon the White Blood Corpuscles." *Research Quarterly Supplement* 6:42-48; May 1935.
93. KEBRIC, BURT M. "Problems of Beginning Teachers of Physical Education in the High Schools of California." *Research Quarterly* 16:42-48; March 1945.

94. KEELER, LINDSEY D. "The Effect of Maturation of Physical Skills as Measured by the Johnson Skill Test." *Research Quarterly* 9:54-58; October 1938.
95. KELLY, CLARENCE E. "An Abstract of an Analysis and Evaluation of Physical Education Activities in the Lafayette, Indiana, Public (Grade) Schools." *Research Quarterly* 12:739-40; December 1941.
96. KELLY, ELLEN DAVIS, and BROWN, JANE E. "The Construction of a Field Hockey Test for Women Physical Education Majors." *Research Quarterly* 23:322-29; October 1952.
97. KENNEY, HAROLD EUGENE. "Evaluative Study of the Required Physical Education Program for Men at the University of Illinois." Doctor's thesis. Urbana: University of Illinois, 1955.
98. KISTLER, J. W. "A Study of the Results of Eight Weeks of Participation in a University Physical Fitness Program for Men." *Research Quarterly* 15:23-28; March 1944.
99. KNAPP, CLYDE, and DROM, BEULAH. "Existing and Desired Physical Education Activities in 126 Illinois Secondary Schools." *Research Quarterly* 22:345-55; October 1951.
100. KRAUS, HANS, and HIRSCHLAND, RUTH P. "Minimum Muscular Fitness Tests in School Children." *Research Quarterly* 25:178-88; May 1954.
101. KRETCHMAR, ROBERT T. "A Questionnaire Study Concerning the Development of Co-Education in College Physical Education." *Research Quarterly* 21:216-29; October 1950.
102. KRUG, EDWARD A. *Curriculum Planning.* New York: Harper & Brothers, 1950.
103. LAMP, NANCY A. "Volleyball Skills of Junior High School Students as a Function of Physical Size and Maturity." *Research Quarterly* 25:189-200; May 1954.
104. LANDERS, JULIA PARTINGTON. "A Score Card for Evaluating Physical Education Programs for Physically Handicapped Children." *Research Quarterly* 16:216-20; October 1945.
105. LANDISS, CARL W. "Influence of Physical Education Activities on Motor Ability and Physical Fitness of Male Freshmen." *Research Quarterly* 26:295-307; October 1955.
106. LANGSTON, DEWEY F. "Standardization of a Volleyball Knowledge Test for College Men Physical Education Majors." *Research Quarterly* 26:60-68; March 1955.
107. LAPORTE, WM. RALPH, CHAIRMAN. "Eighth Annual Report of the Committee on Curriculum Research." *Research Quarterly* 7:99-128; October 1936.
108. LAPORTE, WM. RALPH. "Ninth Annual Report of the Committee on Curriculum Research of the College Physical Education Association." *Research Quarterly* 7:87-100; March 1937.
109. LAPORTE, WM. RALPH, and BROWNELL, C. L. "Report of the Committee on Curriculum Research." *Research Quarterly* 4:145-61; March 1933.
110. LAPP, V. W. "Pupil Objectives in High School Physical Education." *Research Quarterly* 4:157-67; May 1933.
111. LEAVITT, NORMA M., and DUNCAN, MARGARET M. "The Status of Intramural Programs for Women." *Research Quarterly* 8:68-79; March 1937.
112. LEHMAN, HARVEY C., and WITTY, PAUL A. *The Psychology of Play Activities.* New York: A. S. Barnes & Co., 1927.
113. McCRAW, L. W., and TOLBERT, J. W. "Sociometric Status and Athletic Ability of Junior High School Boys." *Research Quarterly* 24:72-80; March 1953.
114. McGRAW, M. G. *Growth: A Study of Johnny and Jimmy.* New York: D. Appleton-Century Co., 1935.
115. MACKENZIE, DONALD H. "Effects of Various Physical Activities on the Physical Fitness of University Men." *Research Quarterly Supplement* 6:125-43; March 1935.
116. MACKEY, HELEN T. "Job Analysis of Women Supervisors of Physical Education in United States Public Schools." *Research Quarterly* 27:32-40; March 1956.

117. MASLEY, JOHN; HAIRABEDIAN, ARA; and DONALDSON, DONALD N. "Weight Training in Relation to Strength, Speed and Co-ordination." *Research Quarterly* 24:308-15; October 1953.
118. MEREDITH, H. W. *The Rhythm of Physical Growth*. Vol. II, No. 3. Iowa City: State University of Iowa Studies in Child Welfare, 1935.
119. MESHIZUKA, TETSUO. "A Program of Professional Training in Physical Education for Colleges and Universities in Japan." Doctor's thesis. Iowa City; State University of Iowa, 1956.
120. MEYER, MARGARET H. "The Effect of Hard Laboratory Exercise on the Total and Differential Leukocyte Count of Young Women." *Research Quarterly* 28:271-78; December 1947.
121. MILLER, JAMES L. "Effect of Instruction on Development of Throwing for Accuracy of First-Grade Children." *Research Quarterly* 28:132-37; May 1957.
122. MILLER, WILMA K. "Achievement Levels in Tennis Knowledge and Skill for Women Physical Education Major Students." *Research Quarterly* 24:81-90; March 1953.
123. MOFFETT, DONOVAN C. "A Survey of Teachers' Needs in Health Education, Physical Education, and Recreation." *Research Quarterly* 20:417-23; December 1949.
124. MOHR, DOROTHY, and VERNIER, ELMON L., EDITORS. *Physical Education for High School Students*. Washington, D. C.: American Association for Health, Physical Education, and Recreation, a department of the National Education Association, 1955. 406 p.
125. MOORE, BEVERLY YOUNG. "The Attitude of College Women Toward Physical Activity as a Means of Recreation." *Research Quarterly* 12:720-25; December 1941.
126. MOORE, H. E. "A Study of the Subjects Taught by 1037 Teachers of Health and Physical Education." *Research Quarterly* 4:29-37; October 1933.
127. MOORE, ROY B. "An Analytical Study of Sex Differences as They Affect the Program of Physical Education." *Research Quarterly* 12:587-608; October 1941.
128. MOREHOUSE, LAURENCE E., and ALOIA, ALEX D. "Changes in Certification Requirements of Physical Education in 13 States Since 1942." *Research Quarterly* 19:276-81; December 1948.
129. MOREHOUSE, LAURENCE E., and MILLER, HENRY G. "Changes in Minimum Certification Requirements of Physical Education Teachers in 19 States Since 1947." *Research Quarterly* 25:75-83; March 1954.
130. MOREHOUSE, LAURENCE E., and SCHAFF, OSCAR. "Prerequisites for Teacher Certification in Physical Education in the United States." *Research Quarterly* 13:286-98; October 1942.
131. MURPHY, LT. THOMAS W., and WICKENS, J. STUART. "Yale Completes One Year of Its Wartime Physical Training Program." *Research Quarterly* 14:333-40; October 1943.
132. NATIONAL COMMITTEE REPORT ON STANDARDS. "National Study of Professional Education in Health and Physical Education." *Research Quarterly* 6:48-68; December 1935.
133. *National Survey of Health and Physical Education in High Schools*. Bloomington: Bureau of Service and Research, School of Health, Physical Education, and Recreation, Indiana University, 1954.
134. NEAL, J. G. "Professional Preparation of Athletic Coaches in the Public Schools of Minnesota." *Research Quarterly* 28:176-78; May 1957.
135. NEILSON, J. P. "A Curriculum for the Professional Preparation of Physical Education Teachers for Secondary School." *Research Quarterly* 2:217-26; March 1931.
136. NEILSON, J. P., CHAIRMAN. "Report of the Committee on Teacher Training in Physical Education in the United States." *Research Quarterly* 4:51-67; March 1933.

137. NEVERS, JOHN E. "The Effects of Physiological Age on Motor Achievement." *Research Quarterly* 19:103-10; May 1948.
138. NEWARK PHYSICAL EDUCATION ASSOCIATION RESEARCH COMMITTEE. "Aims and Objectives of Physical Education Activities." *Research Quarterly* 8:103-22; March 1937.
139. NEW YORK STATE EDUCATION DEPARTMENT. *What We Like To Do: A Report by Fourth, Fifth and Sixth Grade Boys and Girls on What Is Important to Them.* Physical Education Standards Project Bulletin 5, Albany: The University of the State of New York, 1954.
140. NORRIS, J. ANNA, and SWEET, DOROTHY C. "A Survey to Determine the Status of Graduate Work in Physical Education." *Research Quarterly* 8:3-10; December 1937.
141. OLDS, L. W. "The Study of the Effects of Competitive Basketball upon the Physical Fitness of High School Boys as Determined by McCurdy-Larson Organic Efficiency Tests." *Research Quarterly* 12:254-65; May 1941.
142. OWENS, BLANCHE ELIZABETH. "A Film on Fundamental Procedures in Physical Education for Elementary School Classroom Teachers." Doctor's thesis. Iowa City: State University of Iowa, 1956.
143. PARSONS, B. NORTH, and HERKIMER, HAROLD W. "Notes on 'D' Classes." *Research Quarterly Supplement* 6:95-101; March 1935.
144. PAYNE, EFFIETEE MARTIN. "Critical Evaluation of Selected Physical Education Films." *Research Quarterly* 24:335-44; October 1953.
145. PEIK, W. E., and FITZGERALD, G. B. "The Education of Men Teachers of Physical Education for Public School Service in Selected Colleges and Universities." *Research Quarterly* 5:18-28; December 1934.
146. PHILLIPS, BYRON M. "Evaluation of Men's Physical Education Service Programs in Higher Education in New York State." *Research Quarterly* 26:185-96; May 1955.
147. PHILLIPS, MARJORIE. "Standardization of a Badminton Knowledge Test for College Women." *Research Quarterly* 17:48-63; March 1946.
148. PORTER, LORENA RAY. "Construction of a Film as an Aid for Teachers of Primary Physical Education." Doctor's thesis. Iowa City: State University of Iowa, 1956.
149. PRESIDENT'S COUNCIL ON YOUTH FITNESS. "The President's Conference on Fitness of American Youth." *Journal of Health-Physical Education-Recreation* 27: 6: 8-10; September 1956.
150. PROSCH, FREDERICK. "The Effect of Exercise on Heart Rate." *Research Quarterly* 3:75-82; December 1932.
151. RARICK, G. LAWRENCE. "College and University Physical Education Programs After One Year of War." *Research Quarterly* 14:167-74; May 1943.
152. RARICK, G. LAWRENCE, and MCKEE, ROBERT. "A Study of Twenty Third-Grade Children Exhibiting Extreme Levels of Achievement on Tests of Motor Proficiency." *Research Quarterly* 20:142-52; May 1949.
153. RATH, EMIL. "Grade Placement of Folk Dances." *Research Quarterly* 10:51-56; March 1939.
154. RATH, EMIL. "A Study of the Effect of Different Physical Education Programs on the Strength Index of Ninth-Grade Boys." *Research Quarterly* 13:169-77; May 1942.
155. REID, ALICE RUTH. "The Contribution of the Freshman Year of Physical Education in a Liberal Arts College for Women to Certain Personality Variables." Doctor's thesis. Iowa City: State University of Iowa, 1955.
156. RUGEN, MABEL E. "Curriculum Content and Suitable Minors." *Research Quarterly* 5:89-100; December 1934.
157. RUSSELL, HELEN L. "Teacher Education in Physical Education with Special References to the Major Programs for Women in Selected State Teachers Colleges." *Research Quarterly* 16:3-13; March 1945.

158. SCHWARTZ, HELEN. "Knowledge and Achievement Tests in Girls Basketball on the Senior High School Level." *Research Quarterly* 8:143-56; March 1937.
159. SCHWENDENER, NORMA. "Educational Theories Applied to Physical Education." *Research Quarterly* 1:124-29; October 1930.
160. SCOTT, M. GLADYS. "Achievement Examinations for Elementary and Intermediate Swimming Classes." *Research Quarterly* 11:10[illegible]-11; May 1940.
161. SCOTT, M. GLADYS. "Achievement Examinations for Elementary and Intermediate Tennis Classes." *Research Quarterly* 12:40-49; March 1941.
162. SCOTT, M. GLADYS. "Achievement Examinations in Badminton." *Research Quarterly* 12:242-53; May 1941.
163. SCOTT, TOM. "A Study of the Content of Professional Courses in Physical Education." *Research Quarterly* 10:78-83; May 1939.
164. SEILS, LEROY G. "The Relationship Between Measures of Physical Growth and Gross Motor Performance of Primary-Grade School Children." *Research Quarterly* 22:244-64; May 1951.
165. SEYMOUR, EMERY W. "Comparative Study of Certain Behavior Characteristics of Participant and Non-Participant Boys in Little League Baseball." *Research Quarterly* 27:338-46; October 1956.
166. SHIRLEY, MABEL J. "Response of the Normal Prepubescent Heart to Exercises of Graded Intensities." *Research Quarterly* 6:104-12; October 1935.
167. SKUBIC, ELVERA. "Emotional Responses of Boys to Little League and Middle League Competitive Baseball." *Research Quarterly* 26:342-52; October 1955.
168. SNELL, CATHERINE. "Physical Education Knowledge Tests." *Research Quarterly* 7:73-82; March 1936.
169. SNELL, CATHERINE. "Physical Education Knowledge Tests." *Research Quarterly* 7:77-91; May 1936.
170. SNELL, CATHERINE. "Physical Education Knowledge Tests Developed by the Department of Physical Education for Women, University of Minnesota." *Research Quarterly* 6:78-94; October 1935.
171. STEINHAUS, A. H. "Chronic Effects of Exercise." *Physiological Review* 13:103-47; January 1933.
172. STRADTMAN, ALAN D., and CURETON, T. K. "A Physical Fitness Knowledge Test for Secondary School Boys and Girls." *Research Quarterly* 21:53-57; March 1950.
173. STREET, CLAUDE W. "A Study of Suitable Majors or Minors That Are Used in Combination with Physical Education." *Research Quarterly* 4:38-50; October 1933.
174. STUHR, ELSIE JACOBSEN. "Interests and Abilities as a Basis for Program Planning." *Research Quarterly* 7:92-98; May 1936.
175. THOMPSON, A. W. "Guiding Principles in Teacher Training from the Viewpoint of a State Director." *Research Quarterly* 3:93-98; May 1932.
176. TUTTLE, W. W. "The Effect of Exercises of Graded Intensity on the Leukocyte Count." *Research Quarterly* 6:37-45; October 1935.
177. TUTTLE, W. W., and WALKER, FRANK H. "The Effect of a Season of Training and Competition on the Response of the Hearts of High School Boys." *Research Quarterly* 11:78-81; December 1940.
178. VAN DALEN, D. B. "A Study of Certain Factors in their Relation to the Play of Children." *Research Quarterly* 18:279-90; December 1947.
179. WAGLOW, I. F., and REHLING, C. H. "A Golf Knowledge Test." *Research Quarterly* 24:463-70; December 1953.
180. WAGLOW, I. F., and STEPHENS, FOY. "A Softball Knowledge Test." *Research Quarterly* 26:234-43; May 1955.
181. WALTERS, C. ETTA. "A Study of the Effects of Prescribed Strenuous Exercises on the Physical Efficiency of Women." *Research Quarterly* 24:102-11; March 1953.

182. WEAR, CARLOS L. "The Evaluation of Attitude Toward Physical Education as an Activity Course." *Research Quarterly* 22:114-26; March 1951.
183. WEAR, CARLOS L. "Construction of Equivalent Forms of an Attitude Scale." *Research Quarterly* 26:113-19; March 1955.
184. WENDLER, ARTHUR J. "An Analysis of the Physical Fitness of Freshmen Male Students at the State University of Iowa." *Research Quarterly* 13:323-27; October 1942.
185. WILBUR, ERNEST A. "A Comparative Study of Physical Fitness Indices as Measured by Two Programs of Physical Education: The Sports Method and the Apparatus Method." *Research Quarterly* 14:326-32; October 1943.
186. WILKIN, BRUCE M. "The Effect of Weight Training on Speed of Movement." *Research Quarterly* 23:361-69; October 1952.
187. WRIGHTSTONE, J. WAYNE. "Evaluating Practices in Health and Physical Education." *Research Quarterly* 7:26-35; March 1936.
188. ZORBAS, WILLIAM S., and KARPOVICH, PETER V. "The Effect of Weight Lifting upon the Speed of Muscular Contractions." *Research Quarterly* 22:145-48; May 1951.

Research and the Recreation Program

RALPH H. JOHNSON

In developing a program of recreation, the researcher is confronted by problems that are similar in nature and scope to those faced in building the curriculum in physical education. Participation in an organized recreation program may be regarded as an educational experience, the program of recreation being analogous to the curriculum of the school. Further, many aspects of program planning in recreation are closely related to school planning. The need for school and community co-operation in the planning of recreational and school activities is evident. The co-operative planning of school and recreation programs and the joint use of facilities for such programs have proven to be successful procedures in the states of California, New York, and Wisconsin and in the cities of Chicago and Milwaukee.

The researcher in recreation utilizes all of the research tools described in Chapters 5 and 7, and all of the research methods discussed in Chapters 9, 13, 14, and 15. The nature of the problem and the purpose of the research determine the tools and the methods employed.

Great concern for research in recreation can be noted today, and both basic research and the application of research findings are urgently needed. Brightbill (7:41-42) states in a recent report that there is a lack of knowledge based upon research and study in the field of recreation, and further emphasizes that "it is essential that the individual engaged in recreation research realize fully that recreation is an identifiable area of living with its own objectives, its own techniques, and its own contribution to enriched living."

Brightbill's recommendations (7:42) for pertinent research include:

> What is the influence of recreation on personality growth? On the learning process? On building good physical and mental health? On developing character and citizenship? On stimulating democratic living? On mitigating the extremes of crime and delinquency? As a means of sustaining morale? In encouraging self-discipline and self-improvement?

Research dealing with the effect of participation in recreation programs upon individuals or groups and upon individuals considered in a natural setting holds much promise. The playground, the craft shop, and the swimming pool provide the setting for research that is related to attitudes and motivation and that measures practical and direct results of the program.

Efforts may be directed to the study of specific outcomes by measuring the results of participation in terms of progress toward the objectives of the program, and by measuring the long-range effects of participation in recreation activities.

Larson (16:129) emphasizes the values to be determined by direct research:

> Under controlled conditions and applying various appropriate methods of research, the physical, social and/or educational values resulting from participation would be determined. The research should be broad and comprehensive and stem from the basic purposes or philosophies of a particular society. Changes in people, as a result of participation, would serve as a basis for establishing inherent values of various activities.

Recreation and sports programs may be studied in operation. For example, Little League Baseball and Junior Football offer excellent possibilities for such studies (30, 32, 33). A recent statement by Wolffe (41:119) illustrates the need for such studies:

> Widespread debate is still going on over the issue of football in lower schools. Just three weeks ago a regional physical education director sounded the challenge, "Show me one shred of evidence that playing football below the senior high school level is harmful to the participants." His statement was made to a

> nationwide assembly of physical education directors during his defense of his own school system's football program in lower schools. No one was in a position to provide the evidence he demanded. None of us as yet has sufficient data to produce authoritative answers, one way or another. Since such programs are in operation, the living material is available for study--indeed, for basic research on a problem that is plaguing countless educators and parents.

Attention may be given to philosophical, historical, sociological, and comparative approaches to problems in recreation (17). Attempts to implement research and to bridge the gap between research and practice are badly needed. To obtain effective results, such attempts should utilize the efforts of workers in the field (18, 20). Pooling of findings and communication between workers, researchers, and others who are concerned with recreation is essential.

Research is important to program development, and the research outlook is needed by those who work in complex practical situations. Recreation workers can study whole situations in actual settings. Barnes (4:2), in discussing the role of the teacher in research activities, makes the following observation which is also applicable to recreation workers:

> Practitioners in education are ideally located for research activities. Their day-by-day work keeps them close to the world of reality. . . . They are intimate with the necessary research "subjects." . . . They live in a place extravagantly supplied with "data." . . . They are products of a broad systematic education designed to equip them with competent bases for thought and judgment; a prerequisite for hypothesis making.

HISTORICAL TRENDS

Even though research in recreation is relatively new, the problems to which research methods have been applied are broad and diversified. Prior to 1950 most research studies in recreation were reported in the *Research Quarterly* (11, 14, 24). Since that time, there have been many listings of such studies, and extensive efforts have been made to co-ordinate such listings.

When the Recreation Division of the American Association for Health, Physical Education, and Recreation was established in 1950, a research subcommittee was formed. Seven areas of concern to research workers were listed: philosophy, organizations and agencies, measurement and evaluation, program, leadership, administration, and the profession.

Summaries of Studies. In 1949, a joint conference of representatives of the California State Recreation Commission and the Department of Physical Education at the University of California at Los Angeles emphasized the need for research on problems related to the recreation program (22). It was pointed out that attention needed to be directed to historical studies (9, 10, 35, 36), the effects of social change on recreation, geriatrics (18), population mobility, church recreation, recreation in therapy, and intercultural and interracial studies. The need for interdepartmental cooperative research, using city departments and other agencies as laboratories, was stressed; and it was recommended that attention be given to applied research, the results of which could be immediately utilized.

In 1949, Weatherford (40) reported on research in recreation that covered a ten-year period. He points out some of the problems and difficulties faced by the researcher and emphasizes particularly the complexity of the activities in recreation, and the variability of activities in reference to socioeconomic structure and other factors. The variables in group dynamics and social control are stressed, and it is pointed out that suitable instruments for studying the recreation needs of groups are not available.

In a later report on research problems in recreation, Weatherford (39) listed the major problems as:

1. Measurement and evaluation
2. Programs and projects
3. Leadership and professional preparation
4. Certification and Civil Service qualifications
5. Municipal Recreation Administration
6. Rural Recreation Administration
7. Recreation areas and facilities.

He suggested research projects related to curriculum and program study in items 1, 2, 3, 5, and 6 listed above.

Ruys (25) emphasizes that the number of inventory-type studies far overshadows the number of the purposes of recreation, and indicates a need for additional historical studies and for the analysis of the scientific aspects of recreation. He recommends that studies completed in related fields and directly concerned with basic principles underlying recreation should be collected and analyzed. Such examples as motivation, goal-seeking, level of aspiration, and biological need for play are listed.

Listings (1) of needed research having application to the recreation program cover recreational interests and needs of high school students, recreation for the aged, sociometric studies of effects of activities on status in the group, and studies relating motivation and skill to participation.

Co-operative Research Efforts. Research studies at the University of Illinois have been planned as a result of interest on the part of professional personnel in Illinois, and have been co-ordinated with the interests of graduate students. In 1956, this planning was reported by Brightbill (8:440) as follows:

> The University of Illinois has been working closely with the research committee of the Illinois Recreation Association for several years in identifying and proposing problems for research and study. The major purpose of these cooperative arrangements is to gear graduate research to help solve major recreation problems of the recreation practitioner and simultaneously provide experience in research for those pursuing advanced degrees in recreation at the University.
>
> Problems are elicited from membership of the Illinois Recreation Association, evaluated, rated in terms of priority need, and then assigned to qualified investigators—if mutually acceptable to both the graduate student and the university authorities.
>
> Over sixty projects have already been suggested by association members. More than ten of these studies have been completed and include material on such matters as co-operation with school district, co-ordination of community recreational services, volunteers, minority problems, financial practices, nomenclature, fringe areas, population trends, public relations, park-schools and public-school camping. Latest projects are a study of the backgrounds of recreation personnel in Illinois, recently completed, and a plan for the registration or certification of recreation personnel in Illinois soon to be completed.
>
> An even more intensive effort to secure co-operation between the Illinois Recreation Association and the University of Illinois on recreation research projects will be made in the future. As Russell Perry, President of the Illinois Recreation Association stated, "Plans call for workshop meetings in which the committee lists the relative importance of subjects, the study of which will be an asset to Illinois recreation. Committee members, representing various community sizes and organizational structures, should be able to provide a wide variety of problems suitable to graduate study."

As a result of the co-operative planning described above, a study was completed which served as the basis for the development of public recreation in St. Charles, Illinois (34). In a second study, the historical development of one of the most publicized recreation programs in the United States (the program at Decatur, Illinois) was examined (28).

At the National Recreation Congress in 1956, it was emphasized that the recreation leader or administrator should make a professional approach to his job. Six ways in which any leader may make a contribution to the research program are listed (38):

1. By identifying and stating problems in the field that require research. Problems that must come from the field if recreation is to progress.

2. By collecting basic data, such as recording behavior relating to recreation experience of participants. Case studies would be invaluable contributions to recreation literature.
3. By reading and interpreting results of research. Many practitioners need to learn how to understand and appreciate research.
4. By using and applying the results of research in recreation and related fields. Recreation needs a "critical" approach.
5. By writing simple reports of agency activity using accepted research procedures. Recreation workers should learn to use the "tools" of the trade—how to state the facts and communicate them to the profession.
6. By co-operating with persons doing research. The resources of recreation agencies, personnel, facility, and program-wise should be made available for research purposes.

Summary. The principal emphasis in recreation research has been on the activity program. Research in process has been directed to the study of professional preparation, history, philosophy, and social relationships. Further, it has been indicated that co-operative studies with other fields, rehabilitation studies, and geriatric studies are needed.

Studies in recreation present unique problems, since participation in the programs is voluntary and the control of variables is difficult. The resulting difficulties are offset to a degree by the fact that observations can be made under normal program conditions. The application of testing procedures developed in the behavioral sciences makes it possible to measure individual traits and changes, and to analyze social change and social interaction.

Studies that require the application of scientific methods to problems at the operating level have been recommended. Some recreation departments and agencies have been doing operational research in efforts to solve practical problems. For example, the District of Columbia Recreation Department now employs a full-time research person. Personnel limitations and a lack of specialized personnel have hampered progress, but it has been shown that many opportunities exist in which recreation employees may co-operate with research persons in practical research situations. Results obtained from co-operative research projects may be applied immediately.

The number of professional groups engaged in promoting and conducting recreational activities, and the limited co-operation among such groups, serve to inhibit the application of research methods to problems in recreation. The need to bring together research and program information and to develop procedures for the exchange and co-ordination of these materials has been emphasized. Co-operation among the American Recreation Society, the National Recreation Association, the American Association for Health, Physical Education, and Recreation, and other groups is necessary.

Attention has been directed to co-operative efforts to develop research listings and to the creation of a central clearinghouse to disseminate research information. Some progress has been made toward co-ordination at the program level and in relating the efforts of university and special agencies to research.

RESEARCH METHODS

The research procedures of the classical sciences have been applied to some problems in recreation. Historical and philosophical studies have been completed, but the primary emphasis has been directed toward case studies, appraisal studies, and survey studies of many types (26, 19, 23). The techniques of observation, interview, questionnaire, and group deliberation have been most frequently used. Such procedures have been criticized, but Brightbill (7:42) says:

> Recreation is and always will be inescapably related to personality growth, . . . although nothing must substitute for objectivity in research in this field, interest and confidence in the potentialities of recreation are essential to intelligent exploration of it.

Emphasis has been placed upon the need to apply scientific methods to existing problems, and it has been pointed out that departments and agencies are involved continuously in studies of participation, costs, and interests. The need for co-operative research and for use of research results by all concerned has been stressed. The team approach may be utilized in attacking problems in recreation. The recreation leader, the educator, the physiologist, and the psychologist may work together as a research team, with the recreation person acting as the leader and the initiator of the study.

Specific Techniques. Research techniques listed by Meyer and Brightbill (21:74-80) include:

1. Conferences and meetings
2. Observations
3. Inspections
4. Inventories
5. Interviews
6. Questionnaires

7. Personal documents
8. Library study
9. Group deliberations
10. Survey committees
11. Public-opinion polls
12. Review of records and reports
13. Appraisal and comparison with national standards

Survey studies in recreation have limitations which should be taken into account when implications are made from the results of such studies. However, such studies represent co-operative efforts to solve recreation problems and are important to the successful operation of community programs. The study of resources and limitations in particular situations has merit in the solution of specific problems at the operational level. Preliminary studies can be helpful in initiating new programs; and periodic studies provide information for possible adjustment in program and offer opportunity to evaluate the status of on-going programs.

Meyer and Brightbill (21:81-85) describe several varieties of surveys and survey sponsorship, and list as the elements in the survey procedure the survey committee, representatives of agencies and interested groups (both lay and professional), and the survey team. This team does the actual work, prepares the report for the approval of the survey committee, and publicizes the findings.

Procedure for Limited Studies. Meyer and Brightbill (21:81-82) describe as follows the procedures used in carrying out *limited studies*:

> The limited recreation study is widely used to secure facts on the more immediate and closely related recreation elements and resources. Generally it seeks to provide information on total population and a breakdown according to age, race, and income. It takes special cognizance of the school population, and if possible determines anticipated growth or decrease of the population, as well as its distribution according to the neighborhoods or districts. It gives the highlights of local government, including its financial powers and status, its history, and an administrative and financial analysis of the several departments which may have an interest in recreation. Information on city planning, housing conditions, square miles of territory, and neighborhood boundaries is part of the report. Delinquency, accident, and health rates are sought as well as other social data which may appear to have some relation to the recreation problem. The number, types, and sizes of outdoor areas and indoor centers, together with information on equipment and apparatus, are listed. The number and types of recreation staff personnel and the definition of their duties and responsibilities are also a traditional and necessary part of the limited study. The program is checked against possibly one hundred or more activities commonly found in community recreation systems, including sports and games, arts and crafts, music, dramatics, social recreation, nature and outing activities, educational and

civic affairs, hobbies, and special events. Information is also assembled on matters relating to state and local recreation legislation.

The study includes a listing of the facilities, program, and constituency of voluntary youth-serving agencies, the programs and facilities for recreation of churches, industries, labor groups, private clubs, and organizations. Finally, the facilities, practices, and operations of commercialized recreation including theaters, bowling alleys, taverns, dance halls, skating rinks, and amusement parks, are tabulated.

After the facts are gathered, they are analyzed. This analysis, correlating the several factors, provides the basis for the major and minor recommendations.

It should be emphasized that the recreation survey involves all the agencies in the community (5, 29). Some of the agencies have administrative involvements and provide facilities, and others have primarily program relationships.

Social and Behavioral Research Techniques. There has been a trend to associate research in recreation with research in the social and behavioral sciences (15) and to adapt the techniques and methods employed in those sciences to research studies in recreation. Ruys (25:9) cautions that the researcher must keep in mind that, in spite of many variables, recreation studies are primarily of "man" and suggests that the researcher not get involved with too much detail and forget that "man" is the essential unit. Research in recreation can be geared to the study of "man" on the spot and the knowledge gained from such studies can assist in understanding the total individual or group in action (6). General methods recommended by Ruys are the historical, philosophical, descriptive, collaborative and integrative, genetic, growth and developmental, and experimental and statistical.

In addition to the techniques described in relation to the survey procedure, the research techniques suggested as having merit are sampling procedures, physiological and psychological tests, sociometric techniques, social indices, and case studies.

One of the major difficulties encountered in conducting research in recreation is that of adapting the studies to accepted research procedures. Many of the problems for which answers are sought do not fit classic research patterns. However, hypotheses can be established which provide the link between practical research problems in recreation and investigations which may lead to new facts or to new generalizations. There are many factors which make research in recreation programs difficult. Among the most important are the problems involved in (a) identifying and controlling variables in the broad setting of recreation programs, (b) relating the cause of change to the effects as measured, (c) matching individuals or situations, and (d) maintaining consistency throughout the research study. In experimental research, groups can be carefully matched, the study controlled, and the results measured. It is seldom possible to employ this procedure in attacking the types of problems facing the recreation researcher in the practical situation.

A number of plans (12, 4) developed in social research are recommended for the recreation researcher.

1. *A research design which involves before and after studies.* This procedure involves a single experimental group which is tested to determine characteristic behavior; the experimental feature is applied, and the group is retested; and the change that has occurred is measured and analyzed. A limited use of this technique has been made in recreation. It is possible for recreation workers to use this design when practical considerations prevent the use of control or comparative groups.
2. *A single-group after-phase design.* This procedure is primarily a descriptive one and is concerned with a report of current conditions or status. It involves the collection of data and the determination of the interrelationships of the characteristics found. Such studies are not experimental or predictive in any sense, but they provide the background for later predictive or evaluative studies. Most simple surveys would fall into this category.
3. *The ex post facto experiment.* This is a variation of the *before and after studies* and involves single-group research. The general pattern proceeds from the past to the present and involves a process of selecting and using information already recorded. By this procedure, accumulated information such as health records, participation records, and anecdotal accounts is utilized. The information can be set into an experimental design effectively, and records can be matched for all items except the experimental or study factor.

As contrasted with the experimental studies involving quantitative data and raw scores from tests, or other material from which direct computations can be made, many of the program studies in recreation involve qualitative data such as ratings, verbal scores, and other estimates. These estimates must be translated into numerical scores or percentages. The qualitative data can be handled in quantitative terms, and can be scaled, classified, summarized, and interpreted. Such data can be considered to be useful if the data can be classified in such a way that it can be used to answer a specific question. This process involves coding and the scaling and weighting of responses and results. Descriptive statistics are useful in classifying and summarizing these data. The use of descriptive data as outlined in these study patterns may permit the researcher to assume that his group represents a whole population and not a sample (31). No generalizations to total populations need be made, but the researcher can infer that results may be true of other groups.

Research Clinics. Barnes (37) in his work with home economics teachers has experimented with a procedure which has merit for the field of recreation. He suggests clinics in community programs to:

1. Develop an understanding of research design
2. Foster skill in individual and group use of research results
3. Provide stimulus for use of the research approach to the solution of on-the-spot recreation problems.

He emphasizes the need to see, understand, use, and criticize research techniques and points out that research methods may be learned through experience and repeated practice. In preparing the Illinois Curriculum Program Reports (13), the assumption was made that teachers can be researchers. Recreation leaders can also use research proceedings to improve individual practice and to join with others to solve problems and thereby improve programs.

This viewpoint advocates a broad base of research and a wider circle of those engaged in research. It goes beyond the piecemeal noncontinuous study which can only consider a small part of the program and is often fragmentary. It is suggested that there is a need for:

1. Beginning research at a limited level but in which many studies are involved.
2. Research by personnel on the job. (The implication is made that there are not enough research specialists, and that they are not in contact with real situations.)
3. Recreation leaders to assume responsibility for simplified studies. (They have partial preparation and can initiate and carry out some projects.)
4. Further training of recreation leaders through in-service clinics.

Suggested areas of emphasis for the beginning researcher are descriptive studies, prototypes of research in society (such as the Gallup poll), surveys of past and present with projections to the future, descriptive studies and status studies for information on which to base experimental study, and experimental studies in which one simple variable is added to predict change and to study extent of change.

Anderson (2) recommends that new procedures and tools should be developed to measure effectiveness of programs. Several outcomes of comprehensive efforts to evaluate recreation programs are suggested:

1. To determine extent to which objectives are accomplished
2. To determine degree to which program meets the needs and desires of the community
3. To measure progress of various phases of the program for long-range planning
4. To provide factual information for fund-raising or public relations
5. To compare program to national standards
6. To provide incentive for employees.

Evaluation of leadership, activities, time participation, areas, facilities, finances, and community organization are needed. Activity evaluation is particularly pertinent and is basic to the success of the program. Appraisal schedules, such as the schedule for the appraisal of community recreation prepared by the National Recreation Association, are particularly useful. Recent efforts to relate recreation research to city planning are significant. Co-operative work of architects, city planners, and recreation personnel has resulted in the preparation of standards and holds promise for inventive research in new areas of facilities and program development.

NEEDED RESEARCH IN RECREATION

In summary, it is emphasized that there is a need to determine the contribution of recreation activities to the physical welfare of different age groups. Physical outcomes should be measured and reported in meaningful terms. Moral and social values of recreation should be studied, and the results obtained should be considered in revising recreation programs. Evaluation of the interests and needs of individuals and groups is desirable, and generalizations should be made on the basis of such evaluation. Interdisciplinary relationships need to be analyzed; and procedures for co-operation among personnel involved in health, physical education, recreation, and city planning need to be studied.

The Research Council (1:56) of the American Association for Health, Physical Education, and Recreation has listed the following pertinent suggestions for research studies in recreation:

1. The development of an instrument which may be used to predict the success of the prospective recreation leader
2. A study of the age-levels of readiness of children for the development of basic recreation skills
3. A study of those leadership techniques which have proved most successful in the conduct of various recreation activities
4. A study to determine the factors which cause children and adults to drop out of recreation activities
5. Origin of established recreational interest
6. Collecting hobbies of the residents of a small community
7. Longitudinal study of effects of recreation
8. Analysis of therapeutic value of recreation
9. Study of motivational factors in sports
10. Longitudinal studies of changes in recreational interests and patterns

A recent emphasis on broad training of recreation leaders, to include cultural and general education and attention to scientific and statistical study, offers encouragement for in-service advancement and research in recreation. As stated by Sapora (27:24), "There is a need for the recreation practitioner and the researcher to join hands more closely. Each can learn from the other. No profession can advance when there is too great a gap between theory and practice."

BIBLIOGRAPHY ON THE RECREATION PROGRAM

1. AMERICAN ASSOCIATION FOR HEALTH, PHYSICAL EDUCATION, AND RECREATION. *Research Underway and Research Needed in Health, Physical Education, and Recreation.* Washington, D. C.: the Association, 1954. p. 56, 57. (out of print)
2. ANDERSON, JACKSON M. "Evaluating Community Recreation." *Journal of Health-Physical Education-Recreation* 24: 3:25-26, 42; May 1953.
3. BALEY, JAMES A. "Recreation and the Aging Process." *Research Quarterly* 26: 1; March 1955.
4. BARNES, FRED P. *Practical Research Processes.* Illinois Curriculum Program. Trial Draft. 1957.
5. BATCHELOR, W. C. *Recreation Survey and Large-Range Plan.* Columbus, Ohio: Council of Social Agencies, 1949.
6. BRADEMAS, D. J. *The Leisure Time Interests of Fifty Delinquent Boys Prior To Commitment to the Illinois State Training School for Boys, St. Charles, Illinois.* Master's thesis. Urbana: University of Illinois, 1955. 145 p.
7. BRIGHTBILL, C. K. "Research in Recreation—A Growing Need." *Physical Educator* 9:41-42; May 1952.
8. BRIGHTBILL, C. K. "Research Reviews and Abstracts." *Recreation.* November 1956.
9. DUNLAP, HELEN L. "Games, Sports, Dancing, and Other Vigorous Recreational Activities and Their Function in Samoan Culture." *Research Quarterly* 22:298; October 1951.
10. FINK, RUTH W. "Recreational Pursuits in the Old South." *Research Quarterly* 23:28; March 1952.

11. GLOSS, G. M. "Bibliography of Master's Theses and Doctoral Studies in the Field of Recreation." *Research Quarterly* 11:150; March 1940.
12. GOODE, WILLIAM J., and HATT, PAUL K. *Methods in Social Research.* New York: McGraw-Hill Book Co., 1952. 386 p.
13. HAND, H. C. *How To Conduct the Participation in Extra-Class Activity Study.* Series A, Number 51, Illinois Secondary Schools Curriculum Program. Bulletin No. 5. May 1949. 66 p.
14. HJELTE, GEORGE. "Research in Recreation." *Research Quarterly* 10:3; March 1939.
15. JAHODA, MARIE; DEUTCH, MORTON; and COOK, STUART W. *Research Methods in Social Relations.* New York: Dryden Press, 1951. 421 p.
16. LARSON, LEONARD A. "Basic Research in Determining the Inherent Worth of Activities." *Proceedings of the College Physical Education Association.* Washington 6, D. C.: American Association for Health, Physical Education, and Recreation, a department of the National Education Association, 1956.
17. LARSON, LEONARD A., and YOCOM, RACHAEL D. *Measurement and Evaluation in Physical, Health, and Recreation Education.* St. Louis: C. V. Mosby Co., 1951. 507 p.
18. LARSON, RUTH. "Bibliography on Curriculum Research." *Research for Curriculum Improvement.* 1957 Yearbook. Washington, D. C.: Association for Supervision and Curriculum Development, a department of the National Education Association, 1957. Appendix A, p. 67.
19. LUNDBERG, G. A.; KOMAROVSKY, M.; and MCINERY, M. A. *Leisure: A Suburban Study.* New York: Columbia University Press, 1934. 396 p.
20. MCKIM, MARGARET J. "Curriculum Research in Historic Perspective." *Research for Curriculum Improvement.* 1957 Yearbook. Washington, D. C.: Association for Supervision and Curriculum Development, a department of the National Education Association, 1957. 350 p. Chapter 2.
21. MEYER, HAROLD D., and BRIGHTBILL, C. K. *Recreation Administration.* New York: Prentice-Hall, 1956. p. 74-85.
22. MILLER, NORMAN. *A Summary of the Conference on Recreation Research.* Los Angeles: California State Recreation Commission and Department of Physical Education, University of California, 1949. 16 p.
23. NATIONAL RECREATION ASSOCIATION. *Schedule for the Appraisal of Community Recreation.* Bulletin No. 167. New York: the Association, 1944. 27 p.
24. NEUMEYER, MARTIN H. "The Field of Research in Recreation." *Research Quarterly* 10:11-19; December 1939.
25. RUYS, JOOP. *Research in Recreation.* Urbana: University of Illinois, 1954. p. 9, 20. (Mimeographed)
26. SAPORA, A. V. *A Recreation Survey of the Rockford, Illinois, Park District.* Winnebago County: Community Welfare Council, 1956. 207 p.
27. SAPORA, A. V. "Recreation Surveys, Studies and Research." *Proceedings of the Recreation Planning Institute.* Great Lakes District Recreation Conference. Cincinnati: National Recreation Association, 1957. p. 4.
28. SCHOLER, ELMER A. *A Short History of the Organized Recreation Movement in Decatur, Illinois.* Master's thesis. Urbana: Recreation, University of Illinois, 1957. p. 159.
29. SCHULTZ, ARTHUR L. *Community Supported Recreation in the St. Paul Area.* New York: Community Research Associates, 1954.
30. SEYMOUR, EMERY W. "Comparative Study of Certain Behavior Characteristics of Participant and Non-Participant Boys in Little League Baseball." *Research Quarterly* 27:338-46; October 1956.
31. SIEGAL, SIDNEY. *Non-Parametric Statistics: For the Behavioral Sciences.* New York: McGraw-Hill Book Co., 1956. 312 p.
32. SKUBIC, ELVERA. "Emotional Responses of Boys to Little League and Middle League Competitive Baseball." *Research Quarterly* 26:342-351; October 1955.

33. SKUBIC, ELVERA. "Studies of Little League and Middle League Baseball." *Research Quarterly* 27:97-100; March 1956.
34. SMITH, REX B. "*A Comprehensive Recreation Plan for St. Charles, Illinois, and Vicinity.*" Master's thesis. Urbana: Physical Education, University of Illinois, 1956.
35. STUMPF, FLORENCE, and COZENS, FREDERICK W. "Some Aspects of the Role of Games, Sports, and Recreational Activities in the Culture of Modern Primitive Peoples. I. The New Zealand Maoris." *Research Quarterly* 18:198; October 1947.
36. STUMPF, FLORENCE, and COZENS, FREDERICK W. "Some Aspects of the Role of Games, Sports and Recreational Activities in the Culture of Modern Primitive Peoples. II. The Fijians." *Research Quarterly* 20:2; March 1949.
37. UNITED STATES DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, OFFICE OF EDUCATION. "Report of Pacific Regional Conference on Home Economics Education." Part II. *Research Clinic for Teacher Educators, Supervisors, and Administrators in Home Economics Education.* Miscellaneous 3531. Washington 25, D. C.: Superintendent of Documents, Government Printing Office, April 1957. 30 p.
38. WASHINGTON STATE PARKS AND RECREATION COMMISSION, RECREATION DIVISION. *Recreation Views.* Olympia, Washington: the Commission, 1956.
39. WEATHERFORD, ALLEN E. "Major Research Problems in Recreation." *Journal of Health-Physical Education-Recreation* 23:10:19; December 1952.
40. WEATHERFORD, ALLEN E. "A Review of Ten Years of Research in Recreation Education." American Recreation Society *Quarterly Bulletin*, December 1949. p. 22.
41. WOLFFE, JOSEPH B. "Future Basic Research Relating Physical Education to Sports Medicine." *Proceedings of the College Physical Education Association.* Washington 6, D. C.: American Association for Health, Physical Education, and Recreation, a department of the National Education Association. 1956.

Research and the Curriculum in Health Education

WALLACE ANN WESLEY

Health education is concerned with the dissemination of knowledge and the development of habits and attitudes that result in improved personal and community health. In the broad sense, health education deals with such topics as growth, nutrition, the prevention and cure of disease, the correction and adjustment of physical defects, mental health, family relations, and the building of a healthful environment.

Health education is a distinct field of study in the manner that English, mathematics, and biology are fields of study. Health education borrows facts from medicine, physiology, history, anthropology, sociology, psychiatry and psychology, child development, education, and a host of other disciplines. The development of a curriculum in health education requires that these facts be synthesized and integrated into an effective, functional health program. To attain this end, the health educator must seek to understand human behavior and to learn why people do or do not employ the health practices that they know will lead to improved health.

The school program of health education is concerned with the instructional materials used in the school and with all of the school activities in which a knowledge of proper health practices, desirable health habits, and favorable attitudes toward healthful living may be acquired.

Conducting, evaluating, reporting, or recommending needed research related to the curriculum in health education presents several difficulties, one of which is determining the meaning of "curriculum." Because the areas of health services, health instruction, and healthful environment overlap, the curriculum in health education is interpreted here in a liberal sense as the totality of factors in the school and community that affect the health behavior of the pupils. Research related to the curriculum in health education is defined here as the systematic study of the conditions under which human behavior occurs, and the investigation of the kinds of conditions in schools that effectively promote certain desired pupil behavior (13).

Much of the research related to the curriculum in health education has been concerned with the testing of health knowledge and with surveying opinions and attitudes toward health concepts. The success of the efforts directed toward building a sound curriculum in health education might be enhanced if an increased share of the total research efforts were invested in studies concerned with the behavior and the health practices that result from existing programs of health education.

Everyday "common-sense" observations and opinions are not to be disregarded, because an analysis of them may contribute much to the understanding of learning and behavior. However, even though such observations and opinions are useful, systematic research is needed to identify the teaching procedures that contribute most to the development of desirable health habits and favorable attitudes toward healthful behavior.

HISTORICAL TRENDS

An abundance of excellent materials based upon experience and opinion is available in the field of health education (1, 2, 4, 7, 9, 17, 103, 104, 105, 107, 116). However, this abundance cannot be matched with curriculum materials based upon research.

As might be expected, the philosophy of early programs of health education paralleled the philosophical trends in medicine. Medicine cured; fear of poor health was the basis for health education.

Earlier generations of students studied such subjects as hygiene, anatomy, and physiology. In *Hygienic Physiology* by Steele (133), published in 1872, students learned that "The skeleton is the image of death. Its unsightly appearance instinctively repels us . . ." This information, negative in its effect upon students, was not related in any way to health practices. Hygiene courses were filled with rules of the "don't" type, and often the material included in the courses was completely unrelated to the pupil's environment.

In Cleveland, Ohio, considerable foresight in planning a health program for the schools was shown when, in 1917, a health survey was conducted and summarized (16).

Bliss (155) reported in 1917 on an experiment in which an attempt was made to determine the effect of open windows in the classroom on the incidence of illness among the pupils. He reported that illnesses occurred more frequently when the windows were kept open than when the windows were kept closed.

Also in 1917, a group of physicians, educators, and public officials formed the Child Health Organization in an attempt to improve the health of the public through education. They believed that the schools were the logical place to teach health (70, 139, 23), and they suggested a positive, rather than a negative, approach to teaching health rules. This group sponsored the publication in 1924 of the first book entitled *Health Education* (104).

In 1925, the American Child Health Association (8) surveyed 86 cities to study the health habits of children. They believed that the curriculum in health in the schools should be based on the health status and practices of the students being taught.

In 1925-26, children in the schools of the State of New York were given health examinations. It was found that dental defects accounted for more than half of the defects reported. These figures were used 20 years later in a comparative study by Maxwell and Brown (32). They found 905 defects per 1,000 students in the 1925 group and 818 in the 1945 group. The over-all death rate was about one-third greater in the 1925 group. More children in the 1945 group had had defects adjusted than in the 1925 group.

Turner (142) in 1926 supervised health instruction offered to two groups of fifth- and sixth-grade pupils. The children in two schools received special instruction on matters pertaining to health, while the children in the other two schools did not. To check the effectiveness of the teaching, all of the students were weighed and measured. The experimental group gained slightly more in weight and considerably more in height than did the control group. Although the usefulness of this experiment can be questioned, the fact that this marks the beginning of research on the results of a program of health instruction is significant to health education.

In 1926, Kaiser and others (71) tried to determine the results of a special 20-week course in nutrition. Again, results were determined by weighing and measuring. The results offered little support for the special nutrition class.

In 1927, Wood and Lerrigo (154) developed a habit-inventory scale. This scale was published in their book *Health Behavior*, which was used by many teachers to determine the content of their health courses.

Another early study dealing with health behavior was carried on by O'Neill and McCormick (114). The authors used the observation and questionnaire method to survey the health habits of 3,512 students. They found that the students' habits, in general, were poor.

The health status of draftees in World Wars I and II served as a strong incentive for increased emphasis on the teaching of health and on research. The mobilization of workers and the rapidly shifting population throughout the country intensified the need for better health practices and protections.

Curriculum surveys of health education requirements in the nation's schools indicate that 75 percent of the states provide some health instruction (51). Thirty-four percent of the states report that they have a teachers guide or course of study to assist local schools in developing their instructional programs. Later surveys of a similar type indicate a continued increase in the number of states providing teacher guides.

On the other hand, a state regulation or guide does not always indicate what is really being accomplished in the state. Various surveys of the status of health instruction over a period of ten years indicate a need to improve the quality of health education (32).

A recent trend in the construction of the curriculum in health education is the utilization of the technique known as "action research." For a full discussion of this technique, see Chapter 13.

In general, teachers have been found to be inadequately trained in health education (46, 47, 48, 156). Research regarding teacher qualifications affects the curriculum in that desired changes in pupil behavior cannot be expected if teachers are inadequately trained to develop and present a defensible curriculum in health education.

The future of health education looks brighter than in the past because some states now include health education as a requirement for certification of all teachers. Special standards have been set up for those teachers who specialize in health teaching (57). The combined thinking of leaders in the United States concerning the needs of all teachers in the field of health are discussed in *Health Education for Prospective Teachers*, a report of the American Association for Health, Physical Education, and Recreation (3).

HUMAN BEHAVIOR RESEARCH TECHNIQUES

In the basic sciences, many of the research techniques employed and some of the results obtained in studies of behavior may be applied to problems in health education.

Much of the available literature concerning the nature of learning applies to all school areas. Blair and others (19) list and discuss much of the research concerned with individual needs, maturation, factors that affect responses, barriers to goal attainment, and other information related to learning. Some specific studies that deal with attitudes, interests, and behavior may be found in the fields of psychology, sociology, and public health (32, 35, 36, 49, 80, 113, 120, 141, 150).

Research in the areas of behavior requires experimental controls to be established that do not influence the behavior normally exhibited by the subjects, a requirement that makes research in this area extremely difficult. Thus far, scales, projective techniques, and inference-opinion checks have not been entirely satisfactory.

One effective type of test that has recently been developed is the life-like situation test. This test provides a systematic means of assessing complex behavior of individuals and groups. Many educators hesitate to use the results of such tests because they believe that accurate measurement of the intangibles involved is impossible. This point of view is understandable. However, these intangibles must be measured if improved procedures for influencing health behavior are to be developed (13).

Work constantly goes on to improve the reliability and the validity of the tests designed to measure the various factors that influence behavior and to determine the relationship of these factors to future behavior. In the area of vocational guidance, considerable success has been experienced in predicting vocational success. The application of the techniques utilized in this process might fruitfully be applied to the prediction of behavior related to healthful living.

TESTS OF HEALTH KNOWLEDGE

A great deal of literature is available that will aid the teacher in building his own knowledge tests to measure the information acquired by the students as a result of instruction received in the health education program (92, 93, 118, 119, 120).

Such tests may be used as comparative instruments. In such instances, one form of the test is given before a unit of instruction is begun and another form, covering the same information as the first test, is given at the end of the unit.

Many health knowledge tests have been published. Some cover only one area of health while others are more general in scope. Many are designed for specific grades because of the variation in reading ability between the grade levels. Teachers should examine the test items to determine whether or not the test is useful for their particular group of students.

A list of published health knowledge tests may be found in *Tests and Measurements in Health and Physical Education*, by McCloy and Young (92: 399-401). Additional materials may be found in other sources (42, 43, 123, 125).

HEALTH STATUS AND THE CURRICULUM

Research on health status that influences the content of the curriculum in health education includes the physical examination (45, 51, 55, 85, 103), surveys of health status, and comparison studies. In addition, some examples of findings that should be considered in determining the effectiveness of the curriculum utilize the various instruments for measuring physical growth and development (84, 92, 93).

The use of such measures as the Wetzel Grid, Pryor Width-Weight Tables, Meredith Physical Growth Records, and others of a similar nature gives pupils an opportunity to check their own progress and to use this information as a basis for further study (84). The study of nutrition and growth logically follows the appraising of one's own growth pattern.

Surveys—such as the accident study of the Metropolitan Life Insurance Company, the recent survey of absences of 13,113 school children in California and of the 7,000 children in Kentucky (32), the study of food habits of Wisconsin children, and the report by the National Safety Council on accidental deaths—should do much in the future to shape health curriculums. Dental research on the control of tooth decay is basic to much of what will be included in the curriculum regarding nutrition, dental care, and fluoridation (30).

A number of evaluation instruments—variously termed check-lists, survey forms, inventory charts, appraisal forms, evaluation guides, and opinion-inference forms—have been developed to determine attitudes, opinions, practices, and other health behavior.

In 1947, Lewis (44, 90) studied the interests of 3,600 pupils in grades 4 through 12. As would be expected, he found that boys and girls differed in interests. Eighty percent of all students were interested in why people did or did not like them. Many of the other interests noted would be worthy of consideration in the building of a health curriculum.

Southworth, Latimer, and Turner (131) found that the scores earned by students on health knowledge tests were not reflected in their statements of the health practices that they followed.

Straus (136) found that approximately four-fifths of the boys and two-thirds of the girls who drank alcoholic beverages in college began the habit in high school.

Kirkendall (76) found that many students lacked authentic sex information. His findings were supported by the students themselves in a study conducted by Benefiel and Zimnavoda (18).

Byrd (29) developed a scale for students to use in checking their own attitudes toward various health items. This scale is not intended for use in grading a student but is to be used as a guide in health teaching. Remmer's opinion poll and other examples of inventories and scales are included in the bibliography at the end of this section (8, 42, 43, 52, 69, 125, 140).

To maintain and to improve health, the community and the school must work together. For examples of studies in which the community and the school have co-operated in using various scales and inventories, see 4, 9, 97, 112, 140, 146. Invariably, community participation in school health projects has helped to improve the health curriculum and the facilities of the school (83, 87).

One of the most ambitious evaluation projects in the field of health education combined several of the appraisal procedures described above in an attempt to measure the results of a long-term demonstration reaching into several states. Through questionnaires, tests, checklists, survey procedures and opinion polls, the persons who conducted the study determined that their "experience-centered" program was bringing about worthwhile results (96).

Since leaders in the fields of public health and industry have had extensive experience in utilizing these appraisal techniques, their guidance would be helpful to those who seek to determine the most effective use of such techniques in the schools. Researchers in public health measure the subject's familiarity with selected health terms as an index to understanding in the field (34, 152).

A diversity of unproved methods and devices are used in the teaching of health. An analysis of the various methods (65) indicates a need for a variety of methods of teaching. Of course, the effectiveness of any teaching method is in part dependent upon the individual teacher using it.

Bryan (24), Humphrey (66), Knight (77), Bond (22), and Strang (135) studied the effectiveness of demonstration, lecture, group-leader, and group-discussion methods. In general, they found that all have some merit. The success of the individual technique varies with the age level of the students, the leader, and the time available. They also found that a discussion-decision type of group participation most frequently motivated people to put into action the things that they had learned.

Existing conditions in the community should determine to a large extent the content of the curriculum in health education. However, some units of instruction are generally considered as necessary for any adequate health program and should be common to all curriculums in health education. Such units have been selected by means of analyzing textbooks (38, 56, 63, 64) and the duties of health educators (117).

NEEDED RESEARCH IN HEALTH EDUCATION

There is a need for additional research in many areas of health education. As pointed out by Strang (100), research is needed to determine facts in the following areas: nutrition, the prevention of disease, sleep and relaxation, the benefits of exercise, growth, methods of obtaining desired information about people, methods of improving health services, and techniques of health instruction. Additional suggestions for further research may be found in the section on school health of the *Yearbook of the Public Health Association* (127).

One of the most pressing needs in health education is the need for a method of measuring the complex outcomes of a total situation.

Specific suggestions for research in health education include the following:

1. The development of improved instruments for determining health practices
2. The relationship between opinion and health practices
3. The utilization of the team approach (public-health, school, and sociological personnel) in determining what kinds of health behavior may best be learned in school and what kinds can best be learned through other community agencies
4. The effect of the overcrowding of schools on the physical, mental. and emotional health of the students
5. The sources from which students obtain the information that determines their health attitudes and practices
6. The relationship between health status and intellectual achievement
7. The observation of health practices
8. Longitudinal research (over several years) based on case studies of individuals who have followed normal growth patterns and have not suffered from serious illness

9. A comparison of the results obtained from various physical measures such as the Wetzel Grid, the Meredith Physical Growth Records, and the Pryor Weight-Width Tables, etc.
10. The determination of the grade levels at which health knowledge can be most effectively taught.

BIBLIOGRAPHY ON THE CURRICULUM IN HEALTH EDUCATION

1. American Association for Health, Physical Education, and Recreation. *Fitness for Secondary School Youth.* (Edited by Karl W. and Carolyn W. Bookwalter.) Washington, D. C.: the Association, a department of the National Education Association, 1956.
2. American Association for Health, Physical Education, and Recreation. *Fit To Teach.* Revised and rewritten. The Third Yearbook. Washington, D. C.: the Association, a department of the National Education Association, 1957.
3. American Association for Health, Physical Education, and Recreation. *Health Education for Prospective Teachers.* Report of the National Conference on College Health Education. Washington, D. C.: the Association, a department of the National Education Association, 1956. (out of print)
4. American Association for Health, Physical Education, and Recreation. "Health Education Project Sketches." An Evaluation Project. *Journal of Health Physical Education-Recreation* 28:4:25-27; April 1957.
5. American Association for Health, Physical Education, and Recreation and American School Health Association, Joint Committee on Terminology in Health Education. "Terminology in School Health Education." *Journal of Health-Physical Education-Recreation* 22:7:14; September 1951.
6. American Association of Colleges for Teacher Education and Others, Cooperative Committee of School Health Education. *Health Needs of School Children—A Report of Problems As Seen by Teachers.* New York: the Association, 1954.
7. American Association of School Administrators. *Health in Schools.* Revised edition. Twentieth Yearbook. Washington, D. C.: the Association, a department of the National Education Association, 1951.
8. American Child Health Association, Research Division. *Health Survey of Eighty-Six Cities.* New York: the Association, 1925.
9. American Dental Association. *A Dental Health Program for Schools.* Chicago: the Association, 1954.
10. American Social Hygiene Association. *Education for Personal and Family Living: A Working Guide for Colleges.* New York: the Association, 1956.
11. American Social Hygiene Association. *Report of the Midwest Project of In-Service Education of Teachers.* New York: the Association, 1956.
12. Arkansas State Department of Education. *School Health Reporting Form.* Little Rock: Division of Instruction, Arkansas State Department of Education, 1955.
13. Association for Supervision and Curriculum Development. *What Shall the High Schools Teach?* 1956 Yearbook. Washington, D. C.: the Association, a department of the National Education Association, 1956.
14. Avery, Curtis E., and Kirkendall, Lester A. *The Oregon Developmental Center Project in Family Life Education.* Portland: E. C. Brown Trust Co., 1955.
15. Avery, Elizabeth S. "Adults Go Back to School for Health." *Journal of Health-Physical Education-Recreation* 24:2:9-10, 58-59; February 1953.
16. Ayres, Leonard P. "The Cleveland School Survey." Summary Volume. Cleveland, Ohio: Survey Committee of the Cleveland Foundation, 1917.
17. Bauer, William W. *Your Health Today.* New York: Harper & Brothers, 1955.

18. BENEFIEL, CALVIN, and ZIMNAVODA, HELEN. "Pupil Reaction to Family Life Education." *California Journal of Secondary Education* 28:363-65; November 1953.
19. BLAIR, GLENN MYERS; JONES, R. STEWART; and SIMPSON, RAY H. *Educational Psychology.* New York: The Macmillan Co. 1955.
20. BLANCHARD, S. "Integration of the School Health Program with Community Health Education." *American Public Health Journal* 26: 625-28; 1936.
21. BLOCK, MARVIN A.; DUTTWEILER, DOROTHEA C.; MILICH, OLGA; POTTER, MILTON G.; and MONNIER, DWIGHT C. "Problems of Alcoholism as Related to Health Education in the Secondary Schools." *Quarterly Journal of Studies on Alcohol* 12:494-516; September 1951.
22. BOND, BETTY WELLS. *Group Discussion-Decision: An Appraisal of Its Use in Health Education.* St. Paul: Minnesota Department of Health, 1956.
23. BRADLEY, G. H. "History of School Health Education in West Virginia 1863-1945." *Research Quarterly* 18:144-51; May 1947.
24. BROWE, JOHN H., and PIERCE, HAROLD. "A Survey of Nutritional Status Among School Children and Their Response to Nutrition Therapy." *Milbank Memorial Fund Quarterly* 28:223-237; July 1950.
25. BRYAN, MAUDE. *Methods and Techniques of Teaching Health.* Master's thesis. San Diego, Calif.: San Diego State College, 1954.
26. BUCK, CARL E. "The Standardization of School Medical Inspection." *American Journal of Public Health* 13:1017-23; 1923.
27. BUCK, N. M., and OJEMANN, R. H. "The Relationship Between Ability in Scientific Thinking and Behavior." *Journal of Experimental Education* 11:415-19; 1942.
28. BUROS, O., EDITOR. *The 4th Mental Measurement Yearbook.* Highland Park, N. J.: The Gryphon Press, 1953.
29. BYRD, OLIVER E. *Byrd Health Attitude Scale.* Stanford, Calif.: Stanford University Press, 1940.
30. BYRD, OLIVER E. *Health Instruction Yearbooks.* Stanford, Calif.: Stanford University Press, 1943-56. (Published annually)
31. BYRD, OLIVER E. "Health Problems of Significance for Course and Curriculum Construction." *Research Quarterly* 21:3-10; March 1950.
32. BYRD, OLIVER E. *School Health Sourcebook.* Stanford, Calif.: Stanford University Press. 1955.
33. CALIFORNIA STATE DEPARTMENT OF EDUCATION. *How's Your School Program?* Bulletin No. 53. Sacramento: the Department, Bureau of Health Education, 1957.
34. CHALL, JEANNES, and DALE, EDGAR. "Familiarity of Selected Health Terms." *Educational Research Bulletin* 29:199-206; November 1950.
35. COCHRAN, WILLIAM G. "Research Techniques in the Study of Human Beings." *Milbank Memorial Fund Quarterly* 33:121-36; 1955.
36. CONTRIL, HADLEY. "The Intensity of an Attitude." *Journal of Abnormal and Social Psychology* 41:129-35; 1946.
37. COOPERATIVE STUDY OF SECONDARY-SCHOOL STANDARDS. "Health and Safety." *Evaluative Criteria.* Washington, D. C.: American Council on Education, 1950.
38. COMMITTEE ON HEALTH CONTENT OF TEXTBOOK. "Accuracy of Health Content of School Textbooks." Yearbook 1952-53. *American Journal of Public Health* 43:128; May 1953.
39. CONNECTICUT STATE DEPARTMENT OF EDUCATION and CONNECTICUT STATE DEPARTMENT OF HEALTH, ADVISORY SCHOOL HEALTH COUNCIL. *Suggested Teacher Reference Guide on the Subject of Fluoridation of Public Water Supplies.* Hartford: the Departments, 1957. (Mimeographed)
40. CROCKER, OLIVE L. "Family Life Education—Some New Findings." *Social Casework* 36:106-13; March 1955.
41. CUSHMAN, WESLEY. "Problem Solving—An Effective Method for Teaching Health." *Journal of School Health* 23:154-59; May 1953.
42. CUSHMAN, WESLEY. *Health Practice Inventory.* Columbus: Ohio State University, 1954. (Unpublished)

43. DEARBORN, TERRY H. *A Checklist for the Survey of Health and Physical Education Programs in Secondary Schools.* Stanford, Calif.: Stanford University Press, 1951.
44. DENVER PUBLIC SCHOOLS. *Health Interests of Children.* Denver, Colo.: Board of Education, 1947.
45. DOWNES, JEAN. "Sickness as an Index of the Need for Health Supervision of the School Child." *American Journal of Public Health* 35:593-601; 1945.
46. DRURY, BLANCHE JESSEN. "Present Practices in Health Instruction in California Secondary Schools." *Research Quarterly* 24:392-400; December 1953.
47. DZENOWAGIS, JOSEPH G., and OTHERS. "Harmful Health and Safety Misconceptions of a Group of Tenth-Grade Girls." *Journal of School Health* 24:240; November 1954.
48. DZENOWAGIS, JOSEPH G.; BAROZONE, JOSEPH; and IRWIN, LESLIE W. "Prevalence of Certain Harmful Health and Safety Misconceptions Among Prospective Elementary Teachers." *Research Quarterly* 26:44-48; March 1955.
49. EISENBERG, P. *Why We Act As We Do.* New York: Alfred A. Knopf, 1947.
50. EPPRIGHT, ERCEL E., and SWANSON, PEARL P. "Distribution of Nutrients Among Meals and Snacks of Iowa School Children." *Journal of the American Dietetic Association* 31:256-60; March 1955.
51. FARLEY, H. KENT. "Status of Statutory Health Instruction in the United States." *Research Quarterly* 21:287-89; October 1950.
52. FAST, CHARLES, and WILLIAMS, SARA. *Health Education Evaluation Forms.* Portland: Oregon Tuberculosis and Health Association, undated.
53. FIBEL, FRIEDRICH. *Parental Opinion Survey of Health Education on the Secondary Public School Level.* Doctor's thesis. Bloomington: University of Indiana, 1954.
54. FIKES, JAMES A. "Health Knowledge of Teachers." *Journal of School Health* 27:247-48; November 1957.
55. GERMAN, JOHN W. "School Health Program Cuts Medical Defects." *Pennsylvania's Health* 15:4-5; April-June 1954.
56. GREENBERG, B. G.; HARRIS, MARY ELLEN; MACKINNON, FRANCES C.; and CHIPMAN, SIDNEY S. "A Method for Evaluating the Effectiveness of Health Education Literature." *American Journal of Public Health* 43:1147-55; 1953.
57. HAAG, JESSIE HELEN. "Health Education as a Requirement for Certification of Secondary School Teachers." *Research Quarterly* 25:286-97; October 1954.
58. HEIN, FRED V. "A Study of Some Means of Presenting Information to Parents About the School Health Program." *Journal of Experimental Education* 15:304-13; June 1947.
59. HICKS, DORA A. "Scope of Health Instruction—Grade 1-12." *Journal of School Health* 23:179-88; June 1953.
60. HILLEBOE, HERMAN E., and OTHERS. "Newburgh-Kingston Caries-Fluorine Study: Final Report." *Journal of the American Medical Association* 52:290-325; March 1956.
61. HIRSH, JOSEPH. "Alcohol Education—Its Needs and Challenges." *American Journal of Public Health* 37:1574-77; December 1947.
62. HOUSTON PUBLIC SCHOOLS. *Health Evaluation Form.* Houston, Texas: Board of Education, undated.
63. HOYMAN, HOWARD S. "Are High School Health Texts Too Difficult?" *Journal of School Health* 25:274-83; December 1955.
64. HOYMAN, HOWARD S. "Brer Rabbit and Our High School Health Texts." *Journal of School Health* 24:286-89; December 1954.
65. HUBBARD, BRUCE L. *Methods for Teaching Health Education.* Doctor's thesis. Stanford, Calif.: Stanford University, 1953.
66. HUMPHREY, JAMES H. "The Place of the Demonstration in Health Education." *Science Teacher* 19:219-20; October 1952.

67. HUTTON, WILLIAM L.; LINSCOTT, BRADLEY W.; and WILLIAMS, DONALD. "Final Report of Local Studies on Water Fluoridation in Brantford." *Canadian Journal of Public Health* 47:89-92; March 1956.
68. JACKSON, C. O "An Evaluation of Health Practices in Interscholastic Athletics for Boys in Illinois." *Research Quarterly* 15:303-309; December 1944.
69. JACKSON, C. O. "Let's Rate Your Health Education Program." *Journal of Health-Physical Education-Recreation* 26:6:29-30; September 1955.
70. JEAN, SALLY LUCAS. "The Past as a Springboard for the Future." *Journal of School Health* 21:1-10; January 1949.
71. KAISER, ALBERT D.; NORTON, HERMAN J.; and WALKER, EDITH M. "The Value of Nutrition Classes to the Undernourished Child; A Comparative Study." *American Journal of Diseases of Children* 31:386-394; 1926.
72. KAPLAN, OSCAR J. "Evaluation of Health Education Activities by Opinion Poll Techniques." *American Journal of Public Health* 41:31-36; August 1951.
73. KELLOGG FOUNDATION. *An Experience in Health Education.* Battle Creek, Mich.: the Foundation, 1950.
74. KILANDER, HOLGER F. "Health Education as a College Entrance Unit." *Journal of School Health* 21:149-60; May 1951.
75. KIRKENDALL, LESTER A. "How Can We Teach About Sex?" *Journal of Health-Physical Education-Recreation* 24:5:15; May 1953.
76. KIRKENDALL, LESTER A. *Sex Concerns of Adolescent Boys.* Portland, Oregon: E. C. Brown Trust Co., September 1952. (Mimeographed)
77. KNIGHT, STANFORD S., and MICHELSON, JOHN M. "Problem vs. Subject." *The Clearing House* 24:3; September 1949.
78. KNUTSON, ANDIE L. "Evaluating Progress in Health Education." *Journal of Health-Physical Education-Recreation* 28:5:21-26; May-June 1957.
79. KNUTSON, ANDIE L. "Pretesting: A Positive Approach to Evaluation." *Public Health Reports* 67:699-703; July 1952.
80. LANGFORD, WILLIAM S.; GILDER, RODMAN, JR.; WILKING, VIRGINIA N.; GENN, MINNIE M.; and SHERRILL, HELEN H. "Pilot Study of Childhood Accidents; Preliminary Report." *Pediatrics* 11; April 1953.
81. LANTAGNE, JOSEPH E. "An Analysis of the Health Interests of 3,000 Secondary School Students." *Research Quarterly* 21:34-39; March 1950.
82. LANTAGNE, JOSEPH E. "Health Interests of 10,000 Secondary School Students." *Research Quarterly* 23:330-46; October 1952.
83. LARRIMORE, EDDA R., and GALIHER, CLAUDIA B. "Effective Health Teaching—A School and Community Project." *Journal of Health-Physical Education-Recreation* 27:8:18-20; November 1956.
84. LARSON, LEONARD A., and YOCUM, RACHAEL DUNAVEN. *Measurement and Evaluation in Physical, Health, and Recreation Education.* St. Louis: C. V. Mosby Co., 1951.
85. LATIMER, J. V. "Health Education Status of High School Pupils and Need for Coordinating School Health Programs." *Journal of School Health* 13:108; 1943.
86. LAWSON, D. E. "Development of Case Study Approaches." *Educational Forum* 16: March 1954.
87. LAYCOCK, S. R. "How the School Can Obtain Community Understanding and Support." *Understanding the Child* 26:20; January 1957.
88. LEE, MARGIE R. "Background Factors Related to Sex Information and Attitudes." *Journal of Educational Psychology* 43:467-85; December 1952.
89. LEWIS, ARTHUR J. "Health Interests of Children." *Health Education Journal* (London) 7:66-73; September 1949.
90. LEWIS, ARTHUR J. "Research Report of Child Health Needs and Interests on Which to Revise Health Instruction in Denver Schools." *Journal of School Health* 17:31-45; February 1947.
91. McCAULEY, H. BERTON; DAVIS, LILLIAN B.; and FRAZIER, TODD M. "Effect on Oral Cleanliness Produced by Dental Health Instruction and Brushing the Teeth in the Classroom." *Journal of School Health* 25:250-54; November 1955.

92. McCloy, Charles Harold, and Young, Norma Dorothy. *Tests and Measurements in Health and Physical Education.* New York: Appleton-Century-Crofts, 1954.
93. Matthews, Donald K. *Measurement in Physical Education.* Philadelphia: W. B. Saunders Co., 1958.
94. Mayshark, Cyrus. "A Health and Safety Attitude Scale for the Seventh Grade." *Research Quarterly* 27:52-59; March 1956.
95. Meir, Frederick A. "Opportunities in General Science for Health Instruction." *Research Quarterly* 22:434-38; December 1951.
96. Michigan Community Health Service Project. *An Evaluation and Interpretation.* Bulletin No. 409. Lansing: Department of Public Instruction, 1946.
97. Michigan Department of Public Instruction. *A Checklist for Surveying the Secondary School Health Program.* Bulletin No. 346. Lansing: the Department, 1946.
98. Moffett, Donovan C. "A Survey of Teaching Needs in Health Education, Physical Education, and Recreation." *Research Quarterly* 20:417-23; December 1949.
99. Mohr, Bertha H. "The School Health Program—A Self Study." *Journal of School Health* 25:283-91; December 1955.
100. Monroe, W. S., editor. "Health Education." *Encyclopedia of Educational Research.* New York: The Macmillan Co., 1950 and 1951.
101. Morsh, Joseph E., and Plenderleith, E. Mavis. "Changing Teachers' Attitudes." *Canadian Journal of Psychology* 3:117-29; September 1949.
102. National Education Association, Educational Policies Commission. *Purposes of Education in American Democracy.* Washington, D. C.: the Commission, 1938.
103. National Education Association and American Medical Association, Joint Committee on Health Problems in Education. *Health Appraisal of School Children.* Second edition. Chicago: American Medical Association, 1957.
104. National Education Association and American Medical Association, Joint Committee on Health Problems in Education. *Health Education.* Fourth edition. Washington, D. C.: National Education Association, 1948.
105. National Education Association and American Medical Association, Joint Committee on Health Problems in Education. *Healthful School Living.* Washington, D. C.: National Education Association, 1957.
106. National Education Association and American Medical Association, Joint Committee on Health Problems in Education. "National Conference for Cooperation in Health Education." *Suggested School Health Policies.* Third edition. Prepared by the National Committee on School Health Policies. Chicago: American Medical Association, 1956.
107. National Education Association and American Medical Association, Joint Committee on Health Problems in Education, *School Health Services.* Washington, D. C.: National Education Association, 1953.
108. National Education Association, Department of Elementary School Principals. *Health in the Elementary School.* Twenty-Ninth Yearbook. Washington, D. C.: the Department, 1950.
109. New York State Council on Health Teaching, Committee on School Health Councils. *Health Instruction.* Bulletin 4. Roslyn, N. Y.: the Council, undated.
110. Nyswander, Dorothy B. *Solving School Health Problems.* New York: The Astoria Demonstration Study, Commonwealth Fund, 1942.
111. Nyswander, Dorothy B. "Some Research Areas in Health Education." *American Journal of Public Health* 43:1016-19; August 1953.
112. Ohio State Department of Education and Ohio State Department of Health. "How's Your School Health Program?" *Ohio's Health.* Columbus: Ohio State Department of Health, August 1953.

113. OJEMANN, RALPH H.; LEVITT, EUGENE E.; LYLE, WILLIAM H., JR.; and WHITESIDE, MAXINE F. "The Effects of a 'Causal' Teacher-Training Program and Certain Curricular Changes on Grade School Children." *Journal of Experimental Education* 24; December 1955.

114. O'NEILL, F. C., and MCCORMICK, MAY G. "Everyday Behavior of Elementary School Children." No. 1057. Albany: University of the State of New York, 1934.

115. PERROTT, GEORGE ST. J. "Physical Status of Young Men, 1918 and 1941." *Milbank Memorial Fund Quarterly* 19:337-344, 1941.

116. RANDALL, HARRIET. *Teacher Health: A Twelve-Year Study of Absences in Excess of Ten Working Days of Certificated Employees in the Los Angeles City Schools.* Los Angeles: Public Schools, Auxiliary Services Division, 1957.

117. RASH, J. KEOGH. "An Evaluation of the Duties of Health Educators." *Journal of School Health* 21:221-27; September 1951.

118. RASH, J. KEOGH. "Refining the Health Education Test—Development and Trial Application." *Journal of School Health* 23:10-14; January 1953.

119. RASH, J. KEOGH. "Refining the Health Education Test: Item Analysis and Standardization." *Journal of School Health* 23:39-43; February 1953.

120. REMMERS, H. H., and GAGE, N. *Educational Measurement and Evaluation.* New York: Harper & Brothers, 1943.

121. REMMERS, H. H.; HORTON, R. E.; and MAINER, R. E. "Attitudes and Knowledge of Youth Related to Health." *Purdue Opinion Report* 13:9, 19; 1954.

122. RICE, THURMAN B., and HEIN, FRED V. *Living.* Revised edition. Chicago: Scott, Foresman & Co., 1954.

123. RUGEN, MABEL E. "Michigan's Community Health Service Project Aimed at Curriculum Improvement in Health Education in Secondary Schools." *Public Health Nursing* 37:448-54; September 1945.

124. RUGEN, MABEL. *Principles for Methods and Materials in Health Education.* Revised edition. Ann Arbor, Michigan: Edwards Letter Shop, 1939. (Mimeographed)

125. RUGEN, MABEL, and NYSWANDER, DOROTHY B. "The Measurement of Understanding in Health Education." *Forty-Fifth Yearbook of the National Society for the Study of Education,* Part I. Chicago: University of Chicago Press, 1946.

126. SCHMIDT, WARREN, and HAIGH, GERRARD. "The Learning of Subject Matter in Teacher-Centered and Group-Centered Classes." *Journal of Educational Psychology* 47; May 1956.

127. SCHOOL HEALTH SECTION. "Research Needs in School Health." Yearbook 1951-52. *Journal of Public Health* 42:133-38; May 1952.

128. SHAW, JOHN H. "Evaluation in the School Health Instruction Program." *American Journal of Public Health* 47:582-87; May 1957.

129. SHAW, JOHN H., and ANDREW, S. C. "A Study of Some Aspects of Evaluation in a Health Education Workshop." *Journal of School Health* 25:95-101; April 1955.

130. SINGER-BROOKS, CHARLOTTE. "Of What Value Are Health Inventories Filled Out by Parents?" *American Journal of Public Health* 42:661-64; June 1952.

131. SOUTHWORTH, WARREN H.; LATIMER, J.; and TURNER, CLAIR E. "Study of the Health Practices, Knowledge, Attitudes, and Interests of Senior High School Pupils." *Research Quarterly* 15:118-36; May 1944.

132. STARR, HELEN. "Today's Pupil: Health-Informed or Health-Educated?" *Journal of Health-Physical Education-Recreation* 25:7:18-22; September 1954.

133. STEELE, JOEL DORMAN. *Hygienic Physiology.* New York: American Book Co., 1872.

134. STRAIN, FRANCES B., and EGGERT, CHESTER L. *Framework for Family Life Education: A Survey of Present Day Activities in Sex Education.* Washington 6, D. C.: American Association for Health, Physical Education, and Recreation, a department of the National Education Association, 1956. 117 p.

135. STRANG, RUTH M., and SMILEY, DEAN F. *The Role of the Teacher in Health Education.* New York: The Macmillan Co., 1946.
136. STRAUS, R., and BACON, S. D. *Drinking in College.* New Haven: Yale University Press, 1953.
137. STREIT, W. K. "Teaching Social Hygiene to Teen-Agers." *Journal of Health-Physical Education-Recreation* 28:2:22-25; February 1957.
138. STRUSSER, HARRY, and SANDLER, HENRY C. "The Dental Health Education and Follow-Up Program in the Elementary and Junior High Schools of New York City." *Journal of the American Dental Association* 42:143-47; February 1951.
139. TEMPKIN, O. "Health Education Through The Ages." *American Journal of Public Health* 30:1091-95; 1940.
140. TEXAS EDUCATION AGENCY. *A Checklist: Appraising the Elementary and Secondary Health Program.* Bulletin 519. Austin: the Agency, 1951.
141. THURSTONE, L. L. "Theory of Attitude Measurement." *Psychological Record* 36:222-41; 1929.
142. TURNER, C. E. "Health Education and Growth." *American Journal of Public Health* 18:1217-30; 1928.
143. UNITED STATES DEPARTMENT OF THE INTERIOR, BUREAU OF EDUCATION. *Index to the Reports of the Commissioner of Education:* 1867-1907. Bulletin H 7. Washington, D. C.: Government Printing Office, 1909.
144. UNITED STATES OFFICE OF EDUCATION, FEDERAL SECURITY AGENCY. *Teacher Education for the Improvement of School Health Programs: A Summary Report of Two Demonstration Workshops—Suggested Standards for the Professional Education of Teachers.* Bulletin No. 16. Washington, D. C.: Superintendent of Documents, Government Printing Office, 1948.
145. UNIVERSITY OF THE STATE OF NEW YORK. *Evaluation of School Health Education.* Albany: New York State Department of Education, 1952.
146. VIRGINIA DEPARTMENT OF EDUCATION. "How Reliable is Health Information?" (Committee Report). *Journal of Health Physical Education-Recreation* 27:4:11; April 1956.
147. WALKER, HERBERT. *Health in the Elementary School.* New York: The Ronald Press Co., 1955.
148. WASHINGTON DISPATCH. *Role Playing in the Classroom: Portfolio of Teaching Techniques.* Washington, D. C.: Educator's Washington Dispatch, 1950.
149. WASHNIK, WALTER M. "What the High School Senior Knows About Health." *Research Quarterly* 28:178-79; May 1957.
150. WATSON, W. S., and HARTMAN, G. W. "Rigidity of a Basic Attitudinal Frame." *Journal of Abnormal and Social Psychology* 34:314-36; 1939.
151. WAYLAND, L. C. N. "Health Education Dramatized." *Health Education Journal,* December 1954.
152. WEXLER, PHILLIP. "Health Educators Can Learn from Advertisers." *Journal of Health-Physical Education-Recreation* 27:7:27-29; October 1956.
153. WHITE, ESTHER BOYD. *A Study of the Influence of the Arkansas Health Education Workshops in the Health Program in the Public Schools in Arkansas.* Master's thesis. Baton Rouge: Louisiana State University, 1955.
154. WOOD, THOS. D., and LERRIGO, M. O. *Health Behavior.* Bloomington, Ill.: Public School Publishing Co., 1927.
155. WOOD, THOS. D., and ROWELL, H. G. "Report of a Controlled Test of Opening Windows in School Classrooms." *Health Supervision and Medical Inspection of Schools.* (Edited by Don C. Bliss.) Philadelphia: W. B. Saunders Co., 1927. p. 351-52.
156. YANCEY, MAUDE S. "Teachers Need Health Facts." *Journal of Health-Physical Education-Recreation* 25:10:14; December 1954.
157. YOUNG, MARJORIE A. C. "A Study of College Health Programs for Prospective Teachers." *American Journal of Public Health* 44:211-16; February 1954.
158. ZELEMY, L. D. "Experimental Appraisal of a Group Learning Plan." *Journal of Educational Research* 34:37-42; 1940.

CHAPTER 13

Action Research

CAMILLE BROWN
RUTH ABERNATHY

ACTION RESEARCH IS A PROCESS WHEREBY INDIVIDUALS OR GROUPS desiring change within a specific situation test the procedures which they feel may result in such change and then, upon arriving at responsibly evaluated conclusions, put these procedures into operation.

The purpose of action research is to change behavior. Whether action research tests a new administrative organization in operation or clarifies the relationship between a selected teaching procedure and a specific need of children, the ultimate goal is focused upon changing the behavior of individuals. Action research seeks to clarify and validate the relationship between a given action and a given goal.

Research procedures utilized in basic and applied research are used in action research. Action research, however, has two unique components: (a) the researchers are the consumers of the research, and (b) the research takes place within the situation where the problem solution is needed and where the results are to be put into operation.

Action research is not merely an action project. It is more than an individual or a group attack on a problem through which individuals may improve their understandings and skills and put these understandings and skills into operation. Action research differs from other research in that it is formalized through the use of research tools and techniques in evaluating and recording the process.

VALUE

In today's world it is imperative to recognize the reality of change. New possibilities for education need to be developed in keeping with changing times. Policies, organizational patterns, administrative details, subject matter, teaching-learning procedures, counseling and guidance tools and techniques, interpretation of school to the community, the professional organization—all are subject to improvement. Change should be in the very fabric of education.

Teachers and leaders are continually involved in making judgments related to change. Judgments should be based upon the most accurate information available and should be tested for applicability. Group discussion is not enough. Change for the sake of change should not be acceptable.

The value of action research should become clear in that it provides an orderly, disciplined base for change. Because responsible evaluation is inherent in the problem solution, results are defensible.

PURPOSE AND MEANING

While the general purpose of action research is to change behavior, at the same time generalizations are arrived at which contribute to the body of knowledge to be further tested. The generalizations are applicable to the same or to a similar situation, but are not universally applicable.

An action research project may fulfill several purposes at once, but it is undertaken with a primary focus. For example, an administrator with leadership responsibilities may participate in action research in order to improve his ways of working with others. The researcher, in such an instance, may be concerned with improving himself in order to improve others and their ways of working. In the process of change in behavior of individuals, administrative procedures and organization may be forced to change to keep pace with the changing functioning of the individuals involved.

On the other hand, the researchers may purposefully use action research procedures in the changing of a structure or an organization to make it possible for others to work or play more effectively. Leaders may relate scientific foundations to the structure and function of the school or recreation program in order to help children, youth, and adults grow in a manner consistent with present social changes. Leaders may need to change themselves in order to be willing to change the school or other institutions to make them consistent with the changing times.

The researchers may be concerned primarily with themselves or with other subjects. In any case, the study is conducted within a situation over which the researchers have some degree of control.

In summary, the purpose may be to change the behavior of the researchers; to change the behavior of others; or to change a framework, an organization, or other structure which may in turn effect changes in the behavior of the researchers or of others.

Research to Change Researchers. The researchers may be parents, children and youth, administrators, teachers and leaders, and other community members, as well as research consultants. Teachers and leaders may be concerned with the conduct of their many committee meetings. They may feel that committee meeting time is not being used profitably. One such group set up a plan for identifying good and poor leadership skills. They first agreed on good leadership skills and then accepted the hypothesis that, if the leadership skills and group member role responsibilities were used as criteria and followed through, the meetings would be better. Each meeting was to be evaluated against the criteria. These individuals were concerned about changes in their own behavior in order to accomplish other goals. All the individuals involved are researchers—they are the subjects.

Research to Change the Behavior of Others. Researchers may be concerned with learning new procedures and their relationship to goals involving the behavior of individuals other than themselves. The goal is to help change the behavior of others while providing researchers with better ways of working. The researchers have a group of subjects within a declared situation. The subjects may not know that they are part of a research project. The researchers are changing something over which they have some degree of control.

Measurement can be made of the changes in behavior of the subjects, or of the possibilities for such changes, if the action hypothesis and goal prove to be closely related. For example, in a playground where the leader was working toward honesty development, she delimited the problem to mean that, when games were played, the individuals of their own volition would identify their own outs or fouls. Her final true action hypothesis was: "If the problem is set in terms of positive action, the pupils will be honest." In dodgeball, the goal became, "Can I immediately respond when I am hit by returning to the circle and quickly being ready to help hit the others in the circle?" The goal had now become, "Am I smart enough to handle a new problem situation?" rather than, "How can I prevent others from knowing I've been hit?" Honesty became redefining the next goal. The teacher learned how to help pupils face new problems rather than cling to the lost goal. Prestige could be had by reaching for the next problem. Pupil behavior changed in this pattern. The teacher learned a new way of working.

The subjects may be pupils in a school or recreation agency, so that the researchers have some control over the situation. In other instances, the subjects may also be members of the research team.

Research to Change the Framework. Researchers may be concerned with changing an administrative organization, a framework, or a structure in an attempt to meet certain goals more efficiently. Structure and function may be so closely related that to change one is to change the other. In fact, the structure may prevent attainment of the desired function.

One problem may be the identification of the relationship between the goal and the organizational pattern. For example, 16-year-old high school girls in a required physical education class in soccer exemplified poor interpersonal relationships. The teacher was concerned about herself as a teacher and tried to change; she used new teaching-learning procedures to bring about the better interpersonal relationships and she studied student health and backgrounds. She finally concluded that the girls had no reason for playing soccer, that they did not wish to play soccer, and that she might improve the interpersonal relationships by changing the structure—i.e., the activity. The hypothesis became, "If the girls are involved in other activities, their interpersonal relationships will change."

PROBLEM AREAS

Some problem areas which relate to the roles of teachers and leaders and which may be dealt with through action research are:

1. Finding better ways to meet needs of personnel. For example, would sharing in certain administrative procedures make for greater happiness on the job?
2. Finding new organizational patterns which better fit the functions —such as testing new patterns of organization for a playground having a changing neighborhood population.
3. Finding better processes for achieving pupil goals while maintaining the status quo within subject matter and school or recreation structure—such as testing ways to help individuals improve interpersonal relationships while working on a nutrition unit, playing softball, or completing a community project.
4. Finding better processes for achieving pupil goals while being willing to change the unit or other structure itself, if necessary, to make for improved interpersonal relationships.
5. Finding better processes for achieving pupil goals by changing the structure and organizational pattern to make them consistent with certain beliefs—such as changing the subject matter organizational pattern of a school to a problem-centered one and testing it in operation; or changing the structure of parks, recreation plans, or even the simple playground; or health programs, physical education programs, or units of work within a program.
6. Finding ways to improve methods of teaching-learning—such as finding out what practices most closely meet the needs, tasks, problems, or objectives of children and youth through health education, physical education, or recreation, or through the interrelationships among the three areas.
7. Finding ways to improve counseling and guidance in health education, physical education, and recreation—such as identifying what practices in these areas most closely fulfill certain health needs of a selected group of young people.
8. Finding procedures which obtain satisfactory results in interpreting the school to the community, such as testing the value of newspapers to parents of a school.
9. Finding ways to improve a professional organization—such as testing ways of improving participation of members.

10. Finding ways to improve the efficiency and effectiveness of group and committee meetings—such as testing leadership skills in staff meetings, PTA meetings, community service committee meetings, youth service meetings, and professional organization meetings.

LEADERSHIP

Research has been primarily the province of the individual project director working alone, or occasionally with one or more research assistants, dedicated to the discovery of new knowledge or the application of identified "truths." The research laboratory has been seen as the location or situation within which variables could be controlled. Sampling has been diligently handled in order that generalizations might be drawn.

Action research, on the other hand, finds its laboratory in the comparatively "uncontrolled" situation of the classroom or the playground or in the organizational framework within which they exist. The sample is the group or situation under investigation; the project director and other researchers are drawn largely from the practitioners involved in the day-to-day conduct of programs.

Action research calls for "creative teamwork" (8: ix). By the very nature of the situation, there is a demand for competent and creative teachers and leaders, as well as for competent research specialists. The teacher or leader is needed by reason of expertness in the given school or playground and his understanding and knowledges, while the research specialist may be needed because of competence in research design and techniques.

Since those involved in the study are also "consumers" of the results, and since the application of results takes place in the situation over which the "consumers" have some degree of control, the role of the teacher or leader is obvious. On the other hand, the role of the consultant research specialist should be equally obvious. The objective point of view, the expertness in analysis, the knowledge of available instrumentalities, the understanding and anticipation of procedural difficulties comprise another area of quality need.

Much of the effectiveness of a given study will depend upon the degree to which the "experts" can work together toward a common goal.

Such co-operative work becomes an enriching experience. The consultants working within the framework of teacher and leader preparation can gain new insights into classroom and playground needs and incorporate them into the preparation of the next "generation" of teachers and leaders. The teacher or leader actively engaged in the classroom or on the playground can gain in confidence by trying new ideas and in competence by making sounder evaluation of procedures used.

Consultant responsibilities go beyond assistance with research and human relations skills, and include the techniques of helping teachers move in the direction of self-initiative and self-direction. Leadership, however, is continuous, with a consultant or status leader assisting in the formulation of the study, the sharp identification of the problem, the appraisal of proposals made, and moving co-operatively toward emergence of leadership within the group. It seems clear that whatever the source, the leader or leaders should be able to envision the total study, understand the techniques needed, strengthen group relationships, and in general serve the group in ways needed to move it toward a satisfactory completion of the project.

STEPS IN ACTION RESEARCH

Identifying the Problem. Whether the study is simple or complex, the initial step is the clear identification of the problem. The concern of the individual or the group is stated. If the action research being planned is a group project, the first statement may seem to be nebulous, and the group may be unable to make a clear statement immediately. At this point it is wise to explore collectively each individual's own concern, and to obtain as thorough an understanding of the situation as possible.

If an individual project is being carried on, the same steps would be used. The individual might even explore his ideas with others to help him formulate his own concern.

Exploring the Concern to Identify the Problem. Each person may state his concern relating to the problem area—why he is concerned and what he thinks his concern is. These ideas are recorded. Each member of the group is charged with a primary responsibility *to listen* in order to be able to identify the real concern of the group, which at times seems to be hidden. After discussing the concerns and hearing the recorded version, the members of the group attempt to state the problem. At this stage, group members may find they are talking about different problems and may wish to divide the initial group into two or more groups, each with a specific problem.

Describing Ideal State of Affairs. After the problem area is identified, the ideal state of affairs relating to the problem should be described—that is, if the problem can be solved, what is the desired solution? If group members are concerned that honesty be developed within the boys and girls with whom they are working, they must first identify what "honesty" means within the situation with which they are concerned. They need to point out the behaviors within this learning situation which exemplify the ideal state of affairs relating to honesty.

The ideal state of affairs is that described by the group for a particular situation. It may not be the same as that described by another group for the same situation or by the same group for a different situation. It is what is best for this situation at this time, as seen by these researchers.

To determine whether the behavior described has the same meaning to all members of the group in actual practice, the group tests the meaning. A systematic plan for identifying the meaning is made and followed by the individuals within the actual situation, identifying the behavior described—for example, honesty. Within the situation, they record behavior examples which are later considered by the whole group. At their next group meeting, the members hear each other and determine whether or not they are all talking about the same thing—that is, does the ideal state of affairs as described in behavior have the same meaning to all members of the group? The group members work and rework until they are able to describe the ideal state of affairs. They help to identify existing bias.

The researchers describe the behavior as seen in the situation over which they have control, rather than the behavior in the lives of the individuals over which they have no control.

Evaluating the Situation. Having described the ideal state of affairs, the researchers are ready to identify the problem or problems by finding out where the situation is in relation to the ideal state of affairs—that is, the researchers evaluate. They measure the situation against the criteria of the ideal state of affairs. The difference between what is desired (the ideal state of affairs) and what the situation is, becomes the identified problem or problems to be solved. The group considers these problems and selects one with which to begin. These first steps may even be considered to be a study. Such study involves relationships and possibilities which may otherwise be overlooked, and limits the area to be studied.

Stating the Action Hypothesis. The action hypothesis is the action suggested to bring about a solution to the problem. The group explores ideas for solving the problem—if we do this, or this, or this, would the problem be solved? What is the most likely action to be selected for problem solution?

Group interaction at this point is important, for ideas may be brought out that one individual alone may not have considered. The group considers the actions suggested and selects the one it believes has the greatest possibility for solving the problem. This is stated as an action hypothesis. That is, if we *do* this, we believe this will bring about a solution to our problem.

When searching for a hypothesis, the researchers should consider many possibilities. The hypothesis to be tested may be found within the feelings of the individuals involved. For example, do the interpersonal relationships within the group make for the attainment of the goal? A change in physical conditions surrounding the experience may contribute to the solution of the problem. The teaching-learning procedures, such as discussion techniques and visual aids, may be considered. Perhaps the hypothesis to be tested is found in the administrative or organizational framework. For example, the subject matter itself may be changed to bring about changes in behavior. These and other possibilities need to be considered in deriving hypotheses to be tested.

Making the Research Design. A plan is made for testing the hypothesis and evaluating the results. The steps to be taken by each member of the group are determined and stated. The tools and techniques needed to determine the relationship of the action hypothesis to the problem are described. All dimensions of the problem are considered, and techniques are described for each dimension. Forms for recording data are described. How the tools are to be administered and how the results will be reported are considered and recorded. The specific information which must be recorded in order to have proper evaluation is determined. Carrying out the action and evaluating the results is seen in its entirety. There is real interdependence between the methods of gathering data and the statistical methods to be used in analyzing them. Channels of communication are considered and planned for.

When appropriate standardized instruments are available, it is wise to use them as one step in achieving valid analyses as well as for comparison purposes. It may be necessary to devise new instruments. If so, they should be pretested to determine whether they are going to measure what they purport to measure. It should also be made possible to consider whether the means for recording data are feasible.

Testing the Action Hypothesis. The plan is carried out within the actual situation where the change is desired by those interested in making the change—the researchers. The relationship of the action taken to the problem is determined through certain evaluative tools previously described. The action takes place, full records are kept, evidence is gathered and recorded, and the data are analyzed.

Deriving Conclusions. The results are described and analyzed. All members participating in the enterprise consider the results, analyze implications, and derive conclusions. They arrive at generalizations which they may wish to retest or put into operation. If the action hypothesis tested does not solve the problem, another hypothesis may be stated and tested. The group may select another problem from those identified and continue toward problem solution.

The selection of evaluative tools or research techniques and the recording and analyzing or data may not be within the skills of the ordinary teacher or recreation leader. The help of skilled consultants—research specialists—may be needed. They contribute skills that the teachers and leaders may not have, and in turn they are aided by the teachers and leaders who contribute different skills.

SUMMARY OF STEPS IN ACTION RESEARCH

Identifying the Problem

1. Consider your situation. Select a problem or problem area about which you, or you and other members of the group, are concerned—an area in which you believe change is needed.
2. Consider whether or not the problem has the same meaning for everyone in the group.
   (a) State the problem or problem area.
   (b) Describe the ideal state of affairs in relation to the problem area.
   - Identify what the group members desire the situation to be.
   - Use examples of concrete evidence to describe what the desired situation should be.
3. Determine what the situation is as related to what is desired by the group. The problem is stated as the difference between what the group or individual desires a situation to be and what the situation actually is.

Stating the Action Hypothesis

1. Consider possible action procedures which may solve the problem.
2. Select the action procedure which is most apt to solve the problem.
3. State the action procedure as an action hypothesis to be tested in the solution of the problem.

Making the Research Design

1. Decide what evidence is needed to determine the degree to which the problem may be solved through the action being taken.
2. Determine the tools and techniques needed to collect the evidence.
3. Plan for recording and treating data.
4. Make a plan for carrying out the action and evaluating the results.

Testing the Action Hypothesis

1. Gather evidence.
2. Record the data.
3. Treat the data.

Deriving Conclusions

1. Analyze the results; determine the relationship between the action hypothesis and the goal or problem.
2. Make generalizations or derive inferences from the evidence.

Setting Next Steps

1. Retest the generalizations resulting from the first action or state a new action hypothesis to be tested.
2. Define a new problem.

BIBLIOGRAPHY

1. Association for Supervision and Curriculum Development, National Education Association. "Encouraging Creativity in Teacher and Learner." *Educational Leadership* 13:270-332; February 1956.
2. Association for Supervision and Curriculum Development, National Educational Association. "Role of Research and Evaluation." *Educational Leadership* 13:398-455; April 1956.
3. Corey, Stephen M. *Action Research To Improve School Practices.* New York: Bureau of Publications, Teachers College, Columbia University, 1953. 161 p.
4. Dunfee, Maxine, and Greenlee, Julian. *Elementary School Science: Research, Theory and Practice.* Washington, D.C.: Association for Supervision and Curriculum Development, a department of the National Education Association. 1957. 67 p.

5. Foshay, Arthur W.; Wann, Kenneth D.; and associates. *Children's Social Values: An Action Research Study.* New York: Bureau of Publications, Teachers College, Columbia University, 1954. 323 p.
6. Jenkins, David H., and Lippitt, Ronald. *Interpersonal Perceptions of Teachers, Students, and Parents.* Washington, D.C.: Division of Adult Education Service, National Education Association, 1951. 119 p.
7. McKim, Margaret E. "Curriculum Research in Historical Perspective." *Research for Curriculum Improvement.* 1957 Yearbook. Washington, D.C.: Association for Supervision and Curriculum Development, a department of the National Education Association, 1957. Chapter 2, p. 14-40.
8. Taba, Hilda. *With Perspective on Human Relations.* Washington, D.C.: American Council on Education, 1955. 155 p.
9. Taba, Hilda, and Noel, Elizabeth. *Action Research: A Case Study.* Washington, D.C.: Association for Supervision and Curriculum Development, a department of the National Education Association, 1957. 58 p.
10. Wann, Kenneth D. "Action Research in Schools." *Review of Educational Research* 23:337-45; October 1953.

CHAPTER 14

The Historical Method

D. B. VAN DALEN

THE HISTORICAL METHOD OF RESEARCH PROVIDES SCHOLARS WITH a tool for securing reliable knowledge about the past. Research workers using this method must collect, classify, and verify facts in accordance with specific standards and must interpret and present the facts in an orderly narrative that will stand the test of critical examination. The same scholarly standards apply whether the problem is concerned with the history of a nation, general education, health, physical education, recreation, or any other area of study. It has been suggested that the historical approach is a method of inquiry that anyone can use who wants to study the past (11:177). Even research workers who do not select historical problems can utilize some techniques of the historical method to evaluate the previous studies relating to their investigations.

PURPOSE AND SCOPE OF HISTORY

Although the roots of historical narration are embedded in the cultural soil of antiquity, the purpose and scope of historical writing have changed through the ages. The first historical accounts were related to the literary arts. They were folk tales and epic poems that sought to entertain, to excite, to inspire. Very early, however, a few ancient Greeks envisioned history somewhat as a science—a search for truth. Thucydides, in the fifth century B.C., aspired to be more than an imaginative story teller. He desired to secure an accurate account of the past that might aid "in the interpretation of the future." He based his writings on his own observations or the reports of eyewitnesses that he subjected to detailed tests of reliability. For many centuries, most historians ignored the research methods and aim of Thucydides. They wrote history with the objective of glorifying the state or church rather than of arriving at objective truth. Historians did not become disciplined by rigorous, critical standards of research until shortly before the twentieth century.

Modern historians generally agree on the techniques to employ in evaluating source materials, but they still argue about the purpose and scope of historical research. Some men are seeking to establish history as a science. Other scholars contend that this transformation can never take place. Members of this latter school of thought argue that history is concerned with a different kind of subject matter than science and therefore requires a different method and interpretation than science.

In general, there is agreement that historians are scientific in certain respects. "The scientific method may be described as consisting of three processes: observation, hypothesis, and experiment" (16:58). Hockett and others contend that modern historians are scientific in that they critically and objectively investigate their source materials and formulate hypotheses—tentative explanations for the occurrence of events or conditions. But, unlike scientists, historians cannot test their hypotheses by direct observation and experimentation—controlled observation. They cannot personally view the health, physical education, and recreation practices of a hundred years ago, nor can they perform experiments that will create conditions exactly as they once were. The events historians study have occurred in the past, and each is unique and nonrepeatable under laboratory conditions. Historians base their work on observations, but usually employ the method of testing the reliability of the reports of observations made by others rather than make the observations themselves. Historians test their hypotheses by re-examining critically the old evidence and searching for fresh information about the past rather than by experimentation and direct observation.

Some historians dispute whether history can be classified as a science on another basis. Science seeks to generalize. Although both scientists and historians may start with propositions about singular, particular, or unique events, the scientists' ultimate objective is to establish broad generalizations—universal laws or theories, such as the theory of relativity and the law of gravitation—that will explain many unrelated, singular events, or conditions. Scientists strive to establish laws that have predictive power—the capacity to predict that certain phenomena presently unknown will eventually appear in specified circumstances.

Some historians believe that constructing laws by generalizing upon historical data is entirely outside their research province. They think that it is their duty to acquire richly detailed knowledge of an event or condition that occurred in a particular time and place in the past and to trace what preceded and succeeded it. They are not concerned about what always, typically, or generally happens; similarities between events; or repeatable aspects of events. They are interested in the unique aspects of a specific event that differentiate it from other events. In their opinion, as soon as a fact becomes merely an instance of a general rule or law, it has lost its identification with the past and, therefore, is no longer a historical fact. Historians of this school show causal relationships between parts of an event or between the conditions existing before and after it, but they do not seek to generalize about the qualities an event has in common with similar events. They do not try to establish generalizations or historical laws that will predict what inevitably will reoccur under certain conditions.

F. M. Fling summarizes the opinion of this school of historians as follows:

> When our attention is directed toward the *uniqueness*, the *individuality* of past social facts, when they interest because of their importance for the *unique evolution of man in his activities as a social being*, in selecting the facts and in grouping them into a complex, evolving whole, we employ the historical method; the result of our work is history.
>
> If, on the contrary, we are interested in *what past social facts have in common*, in the way in which *social facts repeat themselves*, if our purpose is to form *generalizations, or laws*, concerning social activities, we employ another logical method, the method of the natural sciences. We select our facts not for their individuality or for the importance of their individuality for a complex whole, but for what each fact has in common with others and *the synthesis is not a complex, unique whole, but a generalization in which no trace of the individuality of the past social fact remains*. The result of our work is sociology, not history. Thus the work of the historian supplements that of the sociologist.

The historian is interested in quality, individuality, uniqueness; the sociologist in quantity, in generalization, in repetition. (9:16-17)

In contrast to Fling, some men contend that historians must go beyond the description and interpretation of particular events in the past. They believe that it is important to study the past for the lessons it teaches, for the broad generalizations or laws that can be derived from a study of historical facts. Like Thucydides, they want to tell "what has happened and will hereafter happen again according to human nature" (29). They believe historians can discover and formulate the fixed laws that govern human events just as scientists have discovered natural laws that govern phenomena in the physical world.

Although many historians are intrigued with the possibility of establishing historical laws, they recognize the difficulties involved in the process. Arthur M. Schlesinger states the case as follows:

> If it be said that the assumption of the reign of law in History is unscientific, who can say that it is more scientific to assume that the development of man as a social being has been casual, fortuitous, uncontrolled by law? . . . For the immediate future . . . the attention addressed to the discovery of historical laws is certain to grow greater. The difficulties that lie athwart the seeker's path might well deter the stoutest heart because of the profuseness and complexity of the data to be analyzed and the impossibility of establishing control conditions. (26:228)

Some scholars who agree that there can be historical generalization caution that it may not be the comprehensive type that possesses the precise, predictive power of the laws of natural science. Thomas Woody states:

> Cause and effect relationships . . . and the prediction of outcome, after the manner of certain other sciences, have been generally thought to elude the historian's method. In this again, however, the question of the precision and the scope of prediction enters. Though the study of trends has made great progress in the past generation, ambitious pretensions to reading of the future and the guidance of politics, which appealed to Thucydides, and to certain modern historians as well, are rarely found. (34:179-80)

Because of the problems involved in trying to establish historical laws, many scholars believe that "their tools are not suited to dealing with problems of that type and have accepted the more modest role of narrators and interpreters of men's doings, leaving to the newer sciences of sociology and psychology the investigation and formulation of the laws that govern them" (17:7). The arguments of the disputants in the battle to determine whether history is a science, or can become one, indicate that the war will probably continue for years to come. For the most part, however, educators are not deeply involved in this controversy. In general, they hold that there is a place for studies of unique, unrepeatable events as well as studies that trace reoccurring factors, cyclical variations, and similarities between events (11:172-73).

PROCEDURES IN HISTORICAL RESEARCH

Several procedures are involved in the historical method of research: selecting and delimiting the problem; collecting and classifying source materials; criticizing source materials; formulating tentative hypotheses to explain events or conditions; and interpreting and presenting the facts or findings. These are not necessarily separate or successive processes. They may be pursued in various orders. Usually there is considerable shifting back and forth between the steps. However, for the sake of convenience and clarity, they will be considered separately in the following discussion.

Selecting and Delimiting the Problem. Earlier chapters in this book discuss the selection and delimitation of the research problem in detail. Since the general considerations of choosing a topic remain the same regardless of the type of problem, it is sufficient here merely to mention representative historical studies and the need for research in the field.

The dissertations listed in the bibliography (2, 3, 4, 6, 7, 8, 12, 13, 14, 21, 22, 24, 25, 27) are examples of health, physical education, and recreation problems that students have investigated. Abundant opportunities for other research in these fields are available. Relatively little work has been done in the past. As Thomas Woody points out, "Institutions, movements, men and women, associated with the development of play and physical education, are waiting for an historic interview" (34:186). Unless the profession soon devotes more attention to historical research, much important source material will be lost permanently to mankind. With a little probing, scholars can find a multiplicity of urgent and worth-while problems to investigate.

Collecting and Classifying Source Materials. To engage in research, historians must secure adequate and accurate information about the past. To obtain information that will enable them to advance knowledge, they explore two types of source materials—

primary and secondary. Primary sources are original materials themselves. Secondary sources are descriptions of primary sources.

Primary Source Materials. Scholars also divide primary source materials into two categories—documents or traditions, and remains or relics. *Documents or traditions* are reports of events made in oral, written, or pictorial form with the conscious intent of transmitting information. When historians use a documentary source—for example, the minutes of an 1885 Harvard Athletic Committee meeting—they do not observe the event personally but rather rely on the reports of firsthand witnesses. *Relics or remains,* the second type of primary source materials, are objects or materials handed down from the past without the specific intent of imparting information. They constitute an unconscious testimony of incidents in the lives of people. Historians actually see or handle relics or remains, such as playthings found by archeologists. However, they cannot observe the ancient games personally and, thus, must interpret how the toys were used.

Examples of documents and traditions are as follows:

1. *Official Records:* federal, state, or local legislative, judicial, or executive documents, such as constitutions, laws, charters, court proceedings and decisions, tax lists, and vital statistics; church records; and health, physical education, or recreation records of federal and state departments, special commissions, professional organizations, school boards, or administrative authorities, such as minutes of meetings, reports of committees, administrative orders or directives, catalogues, surveys, annual reports, budgets, courses of study, class schedules, salary lists, honors and awards, attendance records, health records, accident reports, and sports records.

2. *Personal Records:* diaries, autobiographies, letters, wills, deeds, contracts, lecture notes, and original drafts of speeches, articles, and books.

3. *Oral Traditions:* myths, folk tales, family stories, dances, games, superstitions, ceremonies, reminiscences of eye witnesses to events, and recordings. (Sometimes these are secondary sources.)

4. *Pictorial Records:* photographs, movies, drawings, paintings, sculpture, and coins.

5. *Published Materials:* newspaper, pamphlet, and periodical articles; literary and philosophical works that convey information about health, physical education, or recreation.

Examples of remains or relics are as follows:

1. *Physical Remains:* buildings, facilities, grounds, furniture, equipment, costumes, implements, awards, and skeletal remains.
2. *Printed Materials:* textbooks, blank diplomas, record blanks, contracts, certificates, attendance forms, report cards, and newspaper advertisements.
3. *Handwritten Materials:* pupil manuscripts, drawings, and exercises.

The preceding classification of primary historical sources is not precise, exclusive, nor complete. The same source material may be either a document or relic depending upon the condition or purpose of its use and the intention of the producer of the original document or relic. For example, a record blank for a track meet is a remain. But, if the record blank is filled in with names of participants, time, and winners, it conveys information intentionally and is a document.

The importance of primary sources cannot be overestimated. They are the basic materials of historical research. "Without them history would be only an empty tale, signifying nothing" (34:185). Therefore, a scholar makes every effort to get as close as possible to the original condition, object, or event he is studying. The original copy of a book is better than a translation; a visit to a playground, stadium, or historical site is better than a picture of it. Examining the remains of a Roman bath is better than reading about it.

A few people have exerted considerable effort in an endeavor to collect and preserve original source materials. A number of the collections, however, are fragmentary and unorganized. Moreover, the locations of many of them are not common knowledge. Health, physical education, and recreation organizations and sports associations have compiled records and statistics concerning their particular areas of interest; some have also collected equipment, costumes, photographs, and other items. For example, materials are located in the Baseball Hall of Fame at Cooperstown, New York; the Basketball Hall of Fame at Springfield, Massachusetts; and the Ski Museum at Oslo, Norway. A few sporting goods companies also have collections of equipment. The American Association for Health, Physical Education, and Recreation in 1934-35 created a Committee for Permanent Historical Records and Exhibits, which became a standing committee in 1937. Since that time, the Association has established a repository for documents and relics at Queens College, New York.

Many historical societies, museums, libraries, high schools, and colleges have preserved materials of interest to scholars. The New York Public Library has an excellent collection of books and pictures on sports and dancing. Records and equipment relating to Dudley A. Sargent are in the gymnasium at Sargent College at Boston University and the Harvard archives. Dr. Frederick W. Luehring's varied physical education collection is located at the University of Pennsylvania. Oberlin College has preserved Dr. Fred E. Leonard's library, which contains many early physical education books.

By talking with "old timers" in the community and the profession; exploring local libraries, old newspapers, second-hand stores, and attics; and visiting sites of playgrounds, gymnasiums, schools, and recreation areas, research scholars can unearth many source materials. They also can find some private collections that are worthy of investigation. For example, Dr. John Neitz of the University of Pittsburgh has an excellent library of old school health textbooks; Major J. F. Leys, 194 Carling Avenue, Ottawa, Canada, has a collection of R. Tait MacKenzie's work; Richard M. Lamb, 2627 Middle Road, Davenport, Iowa, has a comprehensive library of books, articles, and statistics relating to football; Jay Wyatt, 2233 West Street, River Grove, Illinois, has a collection of football rule books dating back to the original edition.[1]

Secondary Source Materials. Secondary source materials differ from primary source materials in that they are not firsthand eyewitness accounts. They are not written by people who have directly observed the event, thing, or condition discussed. Secondary source materials are summaries of information collected by others. Examples of secondary sources are: encyclopedias, almanacs, textbooks, and bibliographies. A source may be secondary or primary depending on its use. For example, a textbook in the history of physical education is usually a secondary source, but it is a primary source for a scholar who is studying how writers organize textbooks in this field.

[1] The author will compile a list of other collections for distribution, if readers will acquaint him with the nature, extent, and location of available source materials.

Primary sources provide the basic materials for historical research, but secondary materials also serve useful purposes. Secondary source materials introduce students to possible problems, inform them of work done by others on their topics, and lead them to primary sources. They give background information to investigators who cannot examine the original sources because of language difficulties, the unavailability of materials, or the lack of training in critical evaluation of specialized materials. Obviously, the worth of a secondary source is directly proportional to the competency of the author and the extent to which he utilizes primary materials. Although secondary source materials are valuable, it is preferable to use the primary source when seeking historical evidence.

Criticizing the Source Materials. Historians are always suspicious of the authenticity and reliability of the raw data they collect from primary and secondary source materials. They realize that "in historical studies doubt is the beginning of wisdom" (19:50), for research based on untrustworthy sources is labor lost. Consequently, scholars do not accept data as facts until after intensively subjecting them to external and internal criticism.

External Criticism. Through external criticism, historians determine the identity and character of the author and the time, place, and circumstances of the document's origination. They try to discover whether a document is an authentic one or a forgery and whether the credited author actually wrote the materials. To analyze the authenticity and origination of source materials, historians examine the text critically and ask some of the following questions: Is the language, style, spelling, handwriting, and printing of the document typical of the author's other work and the period in which it was written? Did the author exhibit ignorance of things a man of his training and time should have known? Did he write about events, things, or places that a man of that period could not have known? Did anyone intentionally or unintentionally alter the manuscript by copying it incorrectly, adding to it or deleting passages? Is this an original draft of the author's work or a copy? If it is a copy, is it reproduced in the exact words of the original? If the manuscript is undated or the author unknown, are there any internal clues in the document that reveal when, where, why, or by whom the document was written?

To engage in external criticism, scholars must have an extensive background in history. In evaluating their source materials, they sometimes also find it necessary to seek help from auxiliary fields, such as philology, chemistry, anthropology, archeology, cartography, numismatics, art, literature, law, paleography, and various modern and ancient languages. Scholars cannot have a knowledge of everything, but they should secure special training in the fields most closely related to their problems. If they cannot undertake the work of textual criticism personally on some points, they must at least seek the opinion of the most competent experts in the field.

Internal Criticism. In external criticism, historians are concerned with the time, place, authorship, and authenticity of the document. In internal criticism, historians are concerned with the *meaning and accuracy of the statements in the document.* In internal criticism they ask two types of questions: What did the author mean by each word and statement? Are the statements the author made accurate and trustworthy?

Investigating the meaning of a statement, technical term, or archaic word can be a very complicated task requiring considerable knowledge of history, laws, customs, and languages. Many words in older documents that are very familiar to us do not have the same meaning today that they had in earlier times. Interpreting words and statements is a less arduous task in more recent publications. Some words in modern usage, however, do not convey the same meaning to all people. For example, when English and American writers use the word "football" or "public school" they are not referring to the same things. In criticizing any document, historians must determine whether the author is writing seriously, humorously, ironically, or symbolically. They must also evaluate whether the writer is voicing his real sentiments or merely pious, polite, or conventional phrases for public consumption. Translated materials naturally must convey the meaning of the original to be of value.

Historians are suspicious of the statements made in source materials until they critically test them. Investigating the accuracy of the materials in a document requires a careful analysis of the author's competency, integrity, prejudices, and self-interests. To evaluate the validity of the author's statements, scholars ask some of the following questions. Is the author accepted as a competent and reliable reporter by other authorities in that special field? Were his facilities, technical training, and location favorable for observing the conditions he reported? Did emotional stress, health conditions, or lack of intelligence cause him to make faulty observations or an inaccurate report? Did he report on direct observations, hearsay, or borrowed source materials? Did he write the document at the time of observation or weeks or years later? Did he write from carefully prepared notes of observations or from memory? Did he have biases concerning any nation, race, religion, person, political party, social or economic group, professional body, period of history, old or new teaching methods, educational philosophy, or activity that influenced his writing? Did anyone financially assist his research work with the hope of securing a report favorable to a specific cause? Did the author write under any economic, political, religious, or social condition that might have caused him to ignore, misinterpret, or misrepresent certain facts? Was he motivated to write by malice, by a desire to justify his acts, or by a desire to win the approval of succeeding generations? Did the author distort or embellish the truth to achieve colorful literary effects? Did the author contradict himself? Are there accounts by other independent, competent observers of different backgrounds that agree with the report of the author?

General Principles of Criticism. In evaluating documents and relics, historians make many judgments. A number of authorities have written excellent discussions on the problems of internal and external criticism (1, 5, 9, 10, 11, 16, 17, 28, 32). Students interested in historical research should consult these and other sources to secure a deeper understanding of this important aspect of their work.

Before initiating private study, students will also find it helpful to review the following principles of criticism suggested by Woody (34:190):

1. Do not read into earlier documents the conceptions of later times.
2. Do not judge an author ignorant of certain events, necessarily, because he fails to mention them (the argument *ex silentio*), or that they did not occur, for the same reason.
3. Underestimating a source is no less an error than overestimating it in the same degree, and there is not more virtue in placing an event too late than in dating it too early by the same number of years or centuries.

4. A single true source may establish the existence of an idea, but other direct, competent, independent witnesses are required to prove the reality of events or objective facts.
5. Identical errors prove the dependence of sources on each other, or a common source.
6. If witnesses contradict each other on a certain point, one or the other may be true, but both may be in error.
7. Direct, competent, independent witnesses who report the same central fact and also many peripheral matters in a casual way may be accepted for the points of their agreement.
8. Official testimony, oral or written, must be compared with unofficial testimony whenever possible, for neither one nor the other is alone sufficient.
9. A document may provide competent and dependable evidence on certain points, yet carry no weight in respect to others it mentions.

Examples of Criticism. Scholars examining health, physical education, and recreation source materials, must exercise the same care as historians in other fields. Statistics always must be questioned. Who collected the statistics on playground participation? Did he figure on a monthly average or a peak day? Did he count once or three times a child who attends morning, afternoon, and evening sessions? Does the fact that a particular health course appears in a college catalogue prove that it was taught or covered the materials listed in the catalogue? Was the author of an article a good friend or a bitter enemy of the man he discusses? Much revered source material must be as carefully criticized as newly discovered documents. For example, the book *Gymnastik für die Jugend* by Johann C. F. Gutsmuth, published in 1793 in Schnepfenthal, Germany, was considerably changed in later editions. When it was translated in England in 1800 and Philadelphia in 1802, it was not only altered and condensed but also attributed to the wrong author.

Outright forgeries of source materials in health, physical education, and recreation do not commonly occur. Some documents, however, have been accepted as reliable for years before someone has submitted them to thorough criticism. For example, in 1907 the Spalding Baseball Commission, which had been appointed to investigate the origination of baseball, reported that Abner Doubleday invented the game at Cooperstown, New York, in 1839. This report was unchallenged by most people for years and was copied in textbooks, newspapers, and sports books.

When Robert W. Henderson later probed this baseball report, he reached some interesting conclusions (15:170-96). The Commission's findings were the work of one member, A. G. Mills, who was a military friend of Doubleday. Mills based his report on a letter written by Abner Graves. No documents by any other person and no contemporary records were presented to support Graves' story. Henderson points out that when Doubleday supposedly originated the game in Cooperstown, he actually was in West Point and did not return to Cooperstown on leave. After retiring from the army, Doubleday wrote many articles for publication but none about baseball. Moreover, when he died in 1893 his obituary notice did not mention that he invented the game.

A critical examination of the Commission's report revealed many other weaknesses. The name "baseball" and some of the rules that Doubleday supposedly invented in 1839 had appeared in print before that time. Although it was claimed that Graves was present when Doubleday traced the first baseball diamond in the dirt, the original Graves' letter did not state this. A later letter that appears to have been written by Graves disclosed that he did not know "where the first game was played according to Doubleday's plan." Moreover, a few books printed before 1839 discussed or illustrated a baseball diamond. Comparisons of the two Graves' letters revealed some inconsistencies. This was not surprising, for the man wrote from memory almost seven decades after the event.

Henderson believes that certain personal factors may have caused members of the Commission to accept the report. Because of the pressure of other duties, they probably did not check the facts very thoroughly. Perhaps patriotic prejudices also influenced their decision. Some men were anxious to prove that baseball was of American rather than British origin. The possibility that General Doubleday, a famous Civil War soldier, invented the great American pastime must have appealed to them.

Nature of Historical Proof. After a careful assessment of the evidence, historians determine its worth. What does it "prove"? Historians are expected to tell the truth—yet what constitutes historical proof? Historians can never be certain that data are absolutely true. There is always the possibility that even the most reliable witness to an event erred in perception or memory. At best, historians can only ascertain a high degree of probability that the data are "true" facts.

Interpreting and Presenting the Facts. Historians do not aimlessly collect source materials, subject them to intensive criticism, and then present the mass of facts—names, events, places, and dates to the public like "beads on a string." Unrelated bits of information do not advance knowledge appreciably. Even if scholars group their facts and arrange their groups in a logical order, they produce a narrative that is little more than a series of disconnected and unexplained events. Isolated facts lack meaning; they "never speak for themselves but only *to* someone who has a hypothesis which he wishes to test" (28:123-24). Consequently, research scholars go beyond the amassing of facts or the mere describing and classifying of them in accordance with their superficial properties. To produce works of value, they formulate tentative hypotheses to explain the occurrence of events and conditions. They seek the underlying patterns or general principles that explain the structural interrelations of the phenomena under study. Having established hypotheses, they search for data to see whether they can confirm them.

In the early stages of research, graduate students usually do not have clearly defined hypotheses. After blocking out an area of study, they explore the literature in a rather general manner for some time. In analyzing their tentative problems, they discover that the data are vague or incomplete, that some elements do not appear to be related to other known elements or to fit into any particular order, or that there are no adequate interpretations for some phenomena. They are puzzled and disturbed! How can they complete the data, systematize the information, and give some interpretation that will explain the unknown factors? Now they stand on the threshold of research! If they can construct hypotheses—explanations—for the unknown phenomena and test them, they may push back the frontiers of knowledge. To build schemes of explanation that account for the factors they are trying to understand, they engage in a high order of conceptualization.

Hypotheses consist of elements expressed in an orderly system of relationships which seek to explain conditions or events that have not yet been verified by facts. Some elements or relationships in hypotheses are known facts and others are conceptual. The conceptual elements are products of research workers' imagination. They leap beyond the known facts to give plausible explanations for unknown conditions. Hypotheses may provide the conceptual elements that complete the known data, conceptual relationships that systematize unordered elements, and conceptual meaning and interpretations that explain the unknown phenomena. Thus, hypotheses logically relate known facts to intelligent guesses about unknown conditions in an effort to extend and enlarge our knowledge (30). Through conceptualization, which makes it possible to introduce elements and relationships that are not directly observable, investigators can go beyond the known data and set up possible solutions to problems.

The explanations or hypotheses proposed by scholars lack proof at the time they construct them. It is their duty to formulate the conceptual and factual elements and relationships in the hypotheses in such a precise and objective manner that they can test the implications of the hypotheses. In the testing process, they painstakingly re-examine old evidence and search for new that will either confirm or disprove the hypotheses.

In constructing and testing hypotheses, historians must be fully aware of their biases. They cannot propose a pet theory and only search for data that supports it. They cannot distort or disregard data in an effort to confirm their hypotheses. To guard against their biases, historians may formulate several hypotheses to explain a particular event or condition. This forces them to make more thorough investigations. They test each of their hypotheses by referring to all of the available data. Whenever they discover any evidence that opposes one proposed explanation of the event or condition, they must reject that hypothesis. By this testing process, they eventually determine which hypothesis best fits all the facts or what modifications are needed in one hypothesis so that it will offer a satisfactory explanation of the facts.

A brief review of the purpose and use of hypotheses reveals that they are excellent synthesizing tools. Hypotheses are tentative principles or generalizations that account for some phenomena. They retain the character of guesses until facts are found to support them. Through appropriate testing situations, the necessary facts are collected. In the conclusion of the study, these findings are organized in terms of the purposes that initiated the investigation. If factual evidence agrees with the original proposals, it confirms the hypotheses; if it disagrees with the original proposals, it discredits the hypotheses. Hence, hypotheses guide investigators in determining what data are relevant and what procedures are appropriate and adequate for testing the suggested solution to the problem. They also provide a framework for stating the conclusions of studies in a meaningful manner. Thus, in the prolonged process of structuring, refining, and testing hypotheses, historians gradually weave masses of facts into complex, causally connected, organic wholes.

After completing their investigations, historians write a well-organized report of their work. A later chapter in this text gives a detailed discussion of the processes involved in reporting research. It is sufficient to state here that investigators' expositions include a statement of the problem, a review of the literature, the basic assumptions underlying the hypotheses, the statement of hypotheses, the methods employed in testing the hypotheses, the findings and conclusions reached, a bibliography, and possibly an appendix. Historians refrain from embellishing their narrations with dramatic flourishes that distort the truth, but they strive for literary excellence. Their objective is to write lucid, lively, logical accounts that are honest and scholarly.

BIBLIOGRAPHY

1. BARNES, HARRY E. *A History of Historical Writing.* Norman: University of Oklahoma Press, 1937.
2. BARTON, HELEN M. *A Study of the Development of Textbooks in Physiology and Hygiene in the United States.* Doctor's thesis. Pittsburgh: University of Pittsburgh, 1942.
3. BENNETT, BRUCE L. *The Life of Dudley A. Sargent, M.D., and His Contribution to Physical Education.* Doctor's thesis. Ann Arbor: University of Michigan, 1947.
4. BRADLEY, GLADYCE. *The History of School Health Education in West Virginia, 1863-1945.* Doctor's thesis. Ann Arbor: University of Michigan, 1946.
5. BRICKMAN, WILLIAM W. *Guide to Research in Educational History.* New York: New York University Bookstore, 1949.
6. DEAN, MARK E. *A History of Intercollegiate Athletics at Indiana State Teachers College.* Doctor's thesis. Bloomington: Indiana University, 1955.
7. DEGROOT, DUDLEY A. *A History of Physical Education in California (1848-1939).* Doctor's thesis. Stanford, Calif.: Stanford University, 1940.
8. EYLER, MARVIN H. *Origins of Some Modern Sports.* Doctor's thesis. Urbana: University of Illinois, 1956.
9. FLING, FRED M. *The Writing of History.* New Haven: Yale University Press, 1920.
10. GARDINER, PATRICK. *The Nature of Historical Explanation.* London: Oxford University Press, 1952.
11. GOOD, CARTER V., and SCATES, DOUGLAS E. *Methods of Research.* New York: Appleton-Century-Crofts, 1954.

12. GREEN, LAWRENCE J. *A Chronology of Changes in Collegiate Football Rules, 1873 to 1954.* Doctor's thesis. Iowa City: State University of Iowa, 1955.
13. HAGELIN, RICHARD H. *A History of First Aid: An Historical Study of the Methods, Materials and Organizations Responsible for Its Conduct.* Doctor's thesis. Ann Arbor: University of Michigan, 1954.
14. HALSEY, ELIZABETH. *The Development of Public Recreation in the Metropolitan Area of Chicago.* Doctor's thesis. Ann Arbor: University of Michigan, 1939.
15. HENDERSON, ROBERT W. *Ball, Bat and Bishop.* New York: Rockport Press, 1947.
16. HOCKETT, HOMER C. *Introduction to Research in American History.* New York: The Macmillan Co., 1932.
17. HOCKETT, HOMER C. *The Critical Method in Historical Research and Writing.* New York: The Macmillan Co., 1955.
18. HOLLIMAN, JENNIE. *American Sports (1785-1835).* Durham, N. C.: Seeman Press, 1931.
19. JOHNSON, ALLEN. *The Historian and Historical Evidence.* New York: Charles Scribner's Sons, 1930.
20. LEONARD, FRED EUGENE, and AFFLECK, GEORGE B. *A Guide to the History of Physical Education.* Philadelphia: Lea & Febiger, 1947.
21. LINDSAY, EDITH. *Origins and Development of the School Health Movement in the United States.* Doctor's thesis. Stanford, Calif.: Stanford University, 1943.
22. MAW, WALLACE H. *Fifty Years of Sex Education in the Public Schools of the United States (1900-1950): A History of Ideas.* Doctor's thesis. Cincinnati: University of Cincinnati, 1953.
23. RICE, EMMETT A.; HUTCHINSON, JOHN L.; and LEE, MABEL. *A Brief History of Physical Education.* Fourth Edition. New York: The Ronald Press Co., 1958.
24. SALLS, DONALD J. *Historical Study of the Physical Education Program in the State Teachers Colleges of Alabama in the Twentieth Century.* Doctor's thesis. New York: New York University, 1955.
25. SAPORA, ALLEN VICTOR H. *The Contributions of Joseph Lee to the Modern Recreation Movement and Related Social Movements in the United States.* Doctor's thesis. Ann Arbor: University of Michigan, 1952.
26. SCHLESINGER, A. M. "History." (Edited by Wilson Gee). *Research in the Social Sciences.* New York: The Macmillan Co., 1929.
27. SMOLENSKY, JACK. *A History of Public Health in Oregon.* Doctor's thesis. Eugene: University of Oregon, 1956.
28. SOCIAL SCIENCE RESEARCH COUNCIL, COMMITTEE ON HISTORIOGRAPHY. *Theory and Practice in Historical Study.* Bulletin 54. New York: the Council, 1946.
29. THUCYDIDES. *The History of the Peloponnesian War.* Translated by H. Dale. London: Bell Co., 1912.
30. VAN DALEN, D. B. "The Function of Hypotheses in Research." *Physical Educator* 26:21-22; March 1957.
31. VAN DALEN, D. B.; MITCHELL, E. D.; and BENNETT, B. L. *A World History of Physical Education.* New York: Prentice-Hall, 1953.
32. VINCENT, JOHN M. *Historical Research.* New York: Henry Holt & Co., 1911.
33. WHITNEY, F. L. *The Elements of Research.* New York: Prentice-Hall, 1946.
34. WOODY, THOMAS. "Of History and Its Method." *Journal of Experimental Education* 15:175-201; March 1947.

(*See also:* "Historical Bibliography." *Journal of Health—Physical Education—Recreation* 31: Part I, April; Part II, May-June 1960.)

CHAPTER 15

Philosophical Methods

ELEANOR METHENY

THERE IS NO SHARP LINE OF DEMARCATION BETWEEN SCIENCE AND philosophy. Both deal with the things and events which man has observed and experienced in the universe in which he exists. Scientists are concerned primarily with finding more precise ways to describe these things and events; the task of the philosopher is to discover the meaning and value of the scientist's facts within the context of man's total comprehension of his own existence.

In practice, science and philosophy are completely interdependent. An eminent philosopher has said, "Science is a necessary pre-condition of philosophy" (15:178) because a philosopher must know the scientist's facts before he can interpret them to discover their meaning and value in relation to the totality of the universe. An equally eminent scientist has noted that the scientist must utilize the philosopher's interpretations and methods to synthesize the "paradoxes, anomalies, and bewilderments" of disparate facts and find an orderly relationship among them before he can establish the context within which meaningful new experiments may be designed (20:24).

The ultimate questions of philosophy relate to the nature of reality, as such. Is "the universe exactly as it appears" to man? Is the universe "ever-changing and evolving"? Is it governed by "certain related, universal, and unchanging laws" (2:173)? As the scientists have continued to add to the store of knowledge about the phenomena of the universe, the philosophers have arrived at different answers to these questions through the years.

As intermediate steps toward finding answers to these large-scale questions, philosophers attempt to establish general principles which seem to account for common characteristics of groups of apparently related facts in such a way as to "eliminate mere arbitrariness" and "satisfy some demand of rationality" (26:142).

At the immediate and practical level, philosophers use these general principles to predict the probable behavior of similar phenomena (things, persons, events) under similar circumstances, and they use these predictions to guide their choices among possible courses of action.

Thus, the philosophical method for approaching a problem involves identification of basic assumptions being made about the universe in which the problem exists; recognition of general principles which provide a rational explanation of the behavior of phenomena within that universe; and interpretation of observations or "facts" about the phenomena in the light of the general principles.

Stated more simply, "Philosophizing is the process of making sense out of experience" (25:270). What shall I do? How shall I do it? What does this mean? Is this desirable? These are all philosophical problems because they involve making decisions or value judgments on the basis of available information interpreted within the scope of general principles which rest on certain basic assumptions. These processes are implicit in every human decision, even though they may not be consciously identified as such. The research worker who is attempting to find defensible answers to questions relating to phenomena, practices, principles, and policies in his own professional area must deliberately seek out and identify these elements in his thinking.

THE METHODS OF PHILOSOPHY

The basic methods of philosophy are logical induction and logical deduction; the tools are analysis and synthesis; and the materials dealt with are the facts which are available. Using the tools of analysis and synthesis, the philosopher works the facts into patterns which identify the relationships among them. Out of these patterned organizations of facts he derives general principles which describe the relationships inherent in the patterns. He states these general principles in the form of hypotheses about these relationships. Then he tests these hypotheses to determine whether the relationships expressed in them do, in fact, exist in the phenomena with which he is dealing.

These steps in the philosophical process can be described and the uses made of them can be illustrated, but there is no formula which tells the philosopher which facts to use, how to arrange them to reveal their relationships, or how to put them together so that the general principle which binds them together is apparent. Logical induction of a general principle from a group of discrete facts is a subjective process. The utilization of this process depends upon the ability of the philosopher to sense these relationships and develop insight into them. Logical deduction is equally subjective. It depends upon the ability of the philosopher to reason logically from the statement of a theory to the consequences of that theory expressed in terms of the behavior of the phenomena to which it is applicable. The personal ability of the philosopher determines the scope of the hypotheses he is able to formulate and test. Few graduate students have the breadth of information and depth of experience which are necessary for developing large-scale hypotheses, but this does not exempt them from the necessity for using the *methods* the philosopher uses. It only limits the size of the problems to which they can successfully apply these methods.

The application of these methods can best be understood by examining their use in relation to a specific problem. But before such an illustration is provided, it is important to emphasize the fact that no research worker, however experienced, has ever yet *proved* that his hypothesis was *true*. The best any research worker can hope for is to demonstrate that his hypothesis is *tenable*, which means that it is logical in the light of the basic assumptions on which it rests, that there is substantial evidence in accord with the theoretical facts which follow as a logical consequence of the hypothesis, and that no known acceptable evidence contradicts these logically derived theoretical facts.

Acceptance of the belief that he cannot *prove* his hypothesis should in no wise discourage the beginning research worker. The history of human knowledge is the history of all of the hypotheses which were once considered tenable but were discredited by observations made by later generations of scientists. This statement may be illustrated, i.e., its tenability as a working hypothesis about the nature of human knowledge may be demonstrated, on a grand scale by noting the many hypotheses which have been considered as tenable explanations of the motion of the stars.

Aristotle (384-322 B.C.) started with the basic assumptions that the earth was fixed, unchanging, and the center of the universe, and that "the circle was the only perfect curve." From these assumptions he developed the hypothesis that all heavenly bodies "must necessarily move in circular orbits" about the earth (8:372).[1] This hypothesis went unchallenged for almost 2,000 years, although the centers of the circles were shifted from time to time by various philosopher-scientists. Ptolemy (127-141 A.D.) described how "the sun, moon and stars . . . revolved around the fixed central earth, while the planets revolved about other centers which themselves revolved around the earth," all in circular orbits (8:371). As the motion of the celestial bodies was more carefully described by astronomers during the next 12 centuries, it was found necessary to modify the Ptolemaic theory by adding a number of circular epicycles to the major cycles to account for these observed movements, and "epicycle was piled on epicycle until the system became exceedingly complex" (8:372).

Eventually Copernicus (1473-1543) hypothesized that "the Ptolemaic system was too complex to be true." He summarized "why the ancients thought the earth was at rest in the middle of the world as its center" (8:56) and then wrote a logical "answer to the aforesaid reasons and their inadequacy" (8:58). He invalidated the basic assumption that the earth was the fixed center of the universe by showing that the planetary motions could be described more simply by assuming that the stationary center of the universe was the sun. But he still had to retain a few epicycles to make his system agree with the facts of observation.

A century later, Kepler (1571-1630) was able to eliminate the epicycles by describing the orbits of the planets as ellipses (8:372). In the following century, Newton (1642-1727) accounted

[1] In these examples, reference is made to works which have been published many times. For the convenience of the student, the references cited are either readily available collections in which excerpts from these works appear, or recent and inexpensive editions of the works themselves. These references are cited in the hope that students may be motivated to read some of these selections, each of which is an outstanding example of the use of philosophical methods of research.

for Kepler's elliptical orbits with his theory of gravitational attraction between masses (8:373). But the final chapter had still not been written. Newtonian physics accounted for the movement of the outer planets, but careful observation of the planets nearest the sun showed that they did not conform to the orbits deducible from the Newtonian theory.

The basic assumption that man's "sun" was more important than all of the other "suns" in the universe had to be abandoned, and man's most basic assumptions about the nature of time and space had to be radically changed before these discrepancies could be accounted for by Einstein's theory of relativity. Had Einstein found the final answer? He did not think so. He left the way open for future discoveries by noting that the theory expressed in his famous equations was applicable only to "a space structure *of the kind described*" (8:482).

Similar examples of the way in which new "facts" upset old basic assumptions (and the theories based upon them) could be drawn from all fields of human knowledge. In biology, Darwin's observations of characteristics of plants and animals led to a new theory of *The Origin of Species* (8:241-71) which challenged long-held assumptions about the nature of man himself. In medicine, Selye (23) philosophized about the symptoms present in all illness, "the syndrome of just being sick," and validated a theory of stress and adaptation which questions many basic assumptions about the nature of illness and the ways in which it may be treated. In education, Dewey (6:449-85) carefully observed certain aspects of human behavior and interpreted them within the philosophical framework of pragmatism to develop new theories about the nature of learning and the nature of the values inherent in educative processes. Yet, even as the tenability of these theories is still being demonstrated by some investigators, other investigators are uncovering new facts and observations which point the way to modification of them.

On the basis of these and countless other examples of the vulnerability of basic assumptions and hypotheses once considered tenable, the honest research worker can only conclude that research has never *proved* anything and in all probability it never will. Research can only demonstrate that within the context of a given set of basic assumptions a stated theory is tenable in the light of the interpretation given to the facts known at that time. Obviously, this statement itself is only a basic assumption, but it is the one which is most tenable in the light of the facts displayed in the history of human knowledge. It is, moreover, a basic assumption which should inspire every research worker to use extreme caution in interpreting the findings of his research and in making statements about the conclusions drawn from them. As Lucretius noted 2,000 years ago in *De Rerum Natura* (On the Nature of Things), men's knowledge of the world they live in has been acquired "by slow degrees as they advanced on the way step by step. Thus, time by degrees brings several things forth before men's eyes, and reason raises it up into the borders of light" (8:40).

But if man can never know "for certain" what facts, theories, and basic assumptions are *true*, how can he know which course of action is *right* when he is confronted with alternative choices? In the area of value judgments, too, the methods of philosophy are the only methods available. Since men cannot wait for some final certainty to guide their choices, they must make their decisions and act on the basis of the most logically valid assumptions, theories, and facts which are available to them in their moment of history. Rational men can make their decisions on the basis of "a systematic and disciplined examination of . . . relevant, available 'facts' and on frames of reference, i.e., the sets of beliefs and assumptions, in and through which these 'facts' are interpreted and processed for choice and action" (4:137). As a result of such a process, they can only decide or act as they would if they knew "for certain" that their conclusions were *true*. But logically, they will always maintain a margin for doubt, being willing to modify their decisions in terms of new facts or theories which may invalidate the facts, the theories, or even the basic assumptions on which their decisions rested.

It is important to recognize that most differences of opinion represent differences in philosophical concepts which underlie the formation of the opinions. These differences may be rooted in acceptance of different basic assumptions, acceptance of different general theories, or differences in the interpretation of the meaning of the same observed facts. The different conclusions about the nature of the universe reached by the great speculative philosophers (7) demonstrate how even the keenest minds may find different meanings and values in the same facts. Reasonably rational men often arrive at illogical conclusions because emotionally they are unable to relinquish beliefs or convictions which satisfy certain of their ego needs. This subtle distortion of reason by emotion has been evidenced in the resistance offered to every new advance in knowledge and to every new theory. But it is on the outcomes of the processes of philosophy, which are always tempered to some extent by man's immediate emotional needs, that all value judgments, decisions, and plans for action finally rest.

APPLICATIONS OF METHODS OF PHILOSOPHY TO RESEARCH

The philosophical processes of reasoning about facts within the framework of basic assumptions which are implemented by general principles is the *sine qua non* of all research studies, no matter what the general design of the study may be, because the three questions implicit in every research study are: What are the facts? What do the facts mean? and What value does this meaning have? In essence, these are all philosophical questions and can be answered only by employing the methods of philosophy at every step in the research study.

For purposes of illustration, a study which emphasizes the accumulation of facts is used, since this is the type of study most frequently undertaken by graduate students. The question asked is: "What are the facts about the relationship of Factor X to some quantitatively measured manifestation of motor performance?" If the study had emphasized *meaning*, the question might be: "What is the significance of motor performance to the performer?" The general procedure would be much the same as in the first study, but the investigator would deal with different *kinds* of facts, and he would need greater insight to guide his analysis, synthesis, logical induction, and logical deduction. If the study deals primarily with *values*, the question might be: "*Should* more time be allotted to motor performance in the school curriculum?" Since the answer to this question would involve decreasing the time allotted to other kinds of performance, it really asks: "What is the relative *value* of motor performance in the total education of a child?" Again the general procedure would follow the same outline, but the investigator would deal with still different kinds of facts and with a wider variety of facts, and he would need both wisdom and experience to guide his subjective use of the processes of philosophy.

The essential steps in the sound development of any research project worthy of the name are listed in the subheads of the discussion which follows. Under each subhead an attempt has been made to show the general application of philosophical method to this step in the research process.

Identifying the Basic Assumptions. Every research study begins with a general question; and every question rests on certain basic assumptions. Unless the investigator identifies these basic assumptions, he will have no basis for *interpreting* the facts he discovers, and accordingly neither his findings nor his conclusions will have *meaning* or *value.*

"How can motor performance be improved?" Some of the basic assumptions implicit in this question are (a) that it is being asked within the context of a universe which is assumed to have certain characteristics; (b) that something called "motor performance" is exhibited by the population of this universe, which is also assumed to have certain characteristics; (c) that the concept "motor performance" can be defined in meaningful terms; (d) that motor performance as so defined can be either quantitatively or qualitatively described in such a way that degrees of difference between two or more samples of it can be detected; (e) that the population has some set of *values within* which the concept "improve" has *meaning;* and (f) that the question asked has *value,* i.e., that it is worth asking because it can possibly be answered and because the answer is worth knowing.

As the investigator identifies these and other basic assumptions he is making, he is able to refine his original question. His assumptions about the nature of the universe and the nature of the human beings who constitute his general population will underlie every aspect of his subsequent thinking and interpretation in ways much too far reaching to be discussed here. Among other things, they will determine his philosophy of education and his concepts of the nature of the learning experience and the values inherent in it. More specifically, he will identify the population to be dealt with in his study as "children attending elementary schools in the United States, who are assumed to be capable of motor performance as subsequently defined." He will discover that he needs to define "motor performance" in terms of specific manifestations of it which can be quantified and measured. He will note that he has assumed that such measures exist and that he can differentiate between two or more such measures. This assumption also seems to imply that equal amounts of differences between these measures always have equal meaning. (This assumption is difficult for him to defend, and he will probably not be able to defend it, but he must recognize the implications it has for the subsequent interpretation of his objective evidence.) He will recognize that he is assuming that these measures of motor performance can be increased or decreased by some factors in the situation which have not yet been identified. He will note that he is using the word "improve" to mean "increase," thus identifying his assumption that "more motor performance" is desirable or valuable for some reason. This will lead him to see that he is asking the question because he hopes to be able to introduce into situations in which his specific population is found certain factors which will increase the amount of their motor performance.

Retracing his steps to (a) weed out the basic assumptions which he cannot defend, (b) eliminate the bias introduced by his personal prejudices, and (c) define the terms he has assumed can be defined, he begins to reduce his question to a form which can be answered by acceptable research procedures.

Defining the Problem. By the subjective process of philosophizing about his own thinking, he has clarified his thinking. Perhaps he defines "motor performance" as "measurements of time and/or space related to running, jumping, and throwing by children of elementary school age." He defines "improvement" as "increase in scores" determined from these measurements of time and/or space. He now knows "what he is talking about," and he can state his question in much more precise form. It may now read: "Are there factors which can be identified in the public school situation of the United States which increase (or decrease) by a measurable amount the scores derived from measurements of time and/or space related to running, jumping, and throwing by children of elementary school age?" This is very cumbersome, of course, but at this point it serves to identify the basic assumptions upon which his study rests; it conceals no hidden *meanings*; and it eliminates the prejudices which were evidence of his own private *value* structure.

Interpreting the Existing Evidence. With the question clearly stated, the investigator can turn to the literature with some confidence because he knows what kind of *facts* he is looking for. When he has assembled as many relevant facts as he can find and has interpreted them in relation to his own basic assumptions, he begins the process of *analysis*. He tries to classify the facts he has found in various ways, guided by "hunches" derived from personal experience with motor performance of children. He groups his facts under various rubrics such as "amount of time given to instruction," "socioeconomic status," "sex," "type of instruction," etc., always looking for threads of relationship which seem to account for common characteristics of groups of facts.

Following these leads which are developed by his own powers of relevant observation, he moves from *analysis* to *synthesis*. He tries to combine his observations about one apparently related group of facts into a clear-cut statement which unites these observations; and he tries to combine smaller groups of facts into meaningful larger groups which are united by some common principle.

As he moves back and forth between analysis and synthesis, the principles or statements of relationships he is looking for must be developed by the method of *logical induction*, because there is no other way to do this. This is a *subjective* process of *thinking about* the information before him. No one can tell him what principle he is seeking. He, himself, does not know what the principle is or will be. The principle is implicit in the data and he must somehow *induce* it to reveal itself. As one philosopher has written, "Most new discoveries are suddenly-seen things that were always there" (16:5). When the investigator "sees" the principle, he will recognize it because it appears to account for certain relationships which seem to be apparent in his diverse data.

Formulating the Hypothesis. Perhaps the principle he has induced is: "The amount of motor performance evidenced by children in the elementary schools is related to the amount of instruction they are given." Recognizing, however, that there are apparently many other factors which seem to be related to the amount of motor performance evidenced by children used as subjects for the studies he has been analyzing, he includes the qualifying phrase "other things being equal." This principle may now be stated as a *hypothesis*.

Designing the Study. The investigator is now ready to design a study which will elicit the kind of facts he needs to *test the hypothesis*. Many different designs will serve, but all of them will rest on his ability to predict or *deduce* the nature of the facts which will most probably or logically be found if his hypothesis is "right." Like logical induction, *logical deduction* is a *subjective* process. It involves weighing all of the available evidence in the light of past experience with similar phenomena and *deciding* what the most *probable* outcomes of a given event will be.

With reference to the present hypothesis, the question is: "*If* the amount of performance evidenced by children in the elementary schools *is* related to the amount of instruction they are given, then *what* facts would I logically expect to find *in actual situations* involving these elements?" The answer, which is not as obvious as it may at first seem, is: "Facts which can be interpreted to mean that children who have had more instruction do manifest more (or less) motor performance than children who have had less instruction." The crucial clause is: "which can be interpreted to mean." There are many *kinds* of facts susceptible to such interpretation. The investigator's decision about the *kind* of facts he will deal with will determine the specific design of his study. A questionnaire study (see Chapters 5 and 9) will elicit either the observations or the opinions of persons experienced in dealing with the phenomena. A descriptive study (see Chapter 9) will assemble facts about what *has* happened. An experimental study (see Chapter 10) will elicit facts about what *does* happen as a result of factors he, himself, introduces into a situation. An action research study (see Chapter 13) will provide facts derived from on-going observation of results obtained by successive introduction of a number of different factors. A study conducted within the confines of a laboratory situation (see Chapter 11) will elicit limited facts about one carefully controlled aspect of the phenomenon he is interested in. The facts derived from each kind of design will need to be *interpreted* in terms of their *meaning* with reference to his specific hypothesis, and the kind of *value* implicit in them.

Whatever design he chooses, he will keep in mind the significance of the clause "other things being equal." Obviously, he can never keep *all* possible factors *equal* in any population, but his previous analysis of the literature will help him determine which factors are *probably* the most significant in influencing the facts he hopes to discover. He must, therefore, design his study in such a way that these factors, at least, are either "controlled" or can be accounted for in his interpretation.

Analyzing the Findings. After he has gathered his data, whatever *kind* they may be, he must return to the process of analysis to identify their apparent relationships. His choice of analytical methods involves philosophical consideration of the applicability of the various types of analysis to his data, the nature of the information each type of analysis can elicit, and the implications of this kind of information in relation to his hypothesis. His questions are: "What is it *possible* to do with data of this type without violating the basic assumptions implicit in either the technique or the data?" "Which is the most *logical* choice among possible alternatives of analytic techniques?"

Eventually, when his analysis is completed, he synthesizes the relationships revealed by the analysis into a series of compact statements, carefully defining each word he uses to assure himself that the meaning he has assigned to it is in accord with his basic assumptions. These statements constitute his *findings of facts as he has interpreted them* by using the methods of philosophy.

Discussion of the Findings. He must now ask another philosophical question: "What do these factual findings *mean* in terms of the total situation to which my hypothesis refers?" Perhaps he has found correlations of .61, .82, and .53 between "amount of instruction" and "measures of motor performance" for six-, seven-, and eight-year-old girls respectively. What is the significance of these *facts* in relation to his hypothesis? How is this significance affected by the *facts* of corresponding correlations of .35, .41, and .50 reported by another investigator? Perhaps he found a difference of .5 seconds between the averages of the time it took for eight- and nine-year old boys to run 40 yards. Or perhaps 40 qualified "experts" said "Yes," and 38 said "No" to the same question. What does this *mean*, i.e. what is the significance of this information?

Testing the Hypothesis. The test of the hypothesis consists of comparing (a) the theoretical facts he has deduced that will logically be expected to exist if his hypothesis is correct and (b) the facts he has elicited in his study and relevant facts reported by other investigators. If the "actual facts" as he has interpreted them are substantially in agreement with the "theoretical facts" which he has logically deduced, then his hypothesis is *tenable* within the limitations imposed by his basic assumptions, the design of his study, the nature of the facts elicited by it, and the extent of the agreement between his logically deduced facts and his "actual facts" as he has interpreted them. Some of his facts may neither agree nor disagree with his hypothetical facts because they are irrelevant to the hypothesis. These "irrelevant" facts may be noted for further investigation, but they do not affect the tenability of his hypothesis. However, if even one well-substantiated relevant fact is contrary to his deduced theoretical facts, he must *reject* his hypothesis as *untenable* in the light of the evidence he has presented.

Stating the Conclusions. If the investigator has been rigorous about utilizing both the concepts and methods of philosophy at every step in his study, there is no uncertainty about the conclusions he may draw. He may conclude: (a) that his hypothesis is *tenable* within the limits noted above; (b) that his hypothesis is *not tenable* within the limits noted above; or (c) that the evidence available is not sufficiently clear-cut to justify conclusions about either the *tenability* or the *non-tenability* of his hypothesis at the present time.

Making Recommendations. The investigator has not *proved* his hypothesis, but if he has demonstrated that it is tenable on the basis of logical reasoning within the limits of his ability to *interpret* factual evidence in relation to *identified basic assumptions* about the nature of the phenomena with which he has been dealing, then it is the "best answer" available at that moment in terms of the *values* identified in his basic assumptions. Accordingly, he may recommend that this conclusion be used as a basis for making decisions regarding actions which will affect the population he has described within the situation he has identified. He may also, by extension, recommend that his conclusion be used as a basis for making decisions in reasonably similar situations involving reasonably similar populations until such time as "better answers" are provided for such situations and populations.

It should be noted, however, that there is no guarantee that his recommendations will be accepted by all persons concerned. If they disagree with his basic assumptions, they must perforce disagree with his conclusions as guides to action. If they disagree with his interpretations and decisions at various points in the study, they will also reject his conclusions on these grounds. The value of having explicitly identified "what he is doing" and "why he is doing it" at each point in the study now becomes apparent. It enables the investigator to determine the *source* of the disagreement expressed by other investigators. A disagreement about *basic assumptions* is almost impossible to resolve, but it must be approached *at the level of basic assumptions.* Differences in *interpretation* are always debatable, but the debate will consist of showing that one interpretation is more *logical* than another in the light of the same basic assumptions. Differences of *observable fact* rest on the way in which the facts were observed and the precision of the observation. But it is fruitless for two investigators to argue about *interpretation* if they are proceeding from different *basic assumptions;* and it is equally fruitless to argue about observed facts if they are *interpreting* those facts in different ways which seem equally logical to them. The investigator who has identified the philosophical foundations upon which his study rests at each step is in a position to identify the *nature* of such disagreements and to approach them rationally, that is to say, logically. The investigator who has failed to identify his philosophical foundations has no *logical* basis for defending the conclusions drawn from his study.

A second set of recommendations may also be made concerning possibly fruitful investigations in the same general area. These recommendations result from insights derived from consideration of many factors noted which were not subjected to test in the present study.

The Advancement of Knowledge. The total of man's knowledge at any time is like an unfinished patchwork quilt, the final plan of which is not known. Each new hypothesis which has been found tenable or untenable in relation to an identified set of basic assumptions on the basis of logical interpretation adds one more patch, even though it may be a small one, which alters the total quilt in some way. The job of the Master's degree candidate is to supply such well substantiated small patches of hypothesis. The doctoral candidate may use a handful of these hypotheses to develop a well-substantiated theory which describes the pattern of some small area in the quilt. The experienced research worker can then synthesize several of these patterns into larger designs of tenable general theories which provide large-scale principles to guide men's actions. And the philosophers will continue to gather these patches, patterns, and designs together and examine them to test the validity of man's basic assumptions about the total plan, the nature of reality, "the intricate web of meaning which is the real fabric of human life" (16:63), and the values man finds in his existence as a human being.

STUDIES IN HEALTH, PHYSICAL EDUCATION, RECREATION

The Concepts of Philosophy. Few studies have been reported in which the concepts which make up the subject matter of philosophy, as such, have been examined with reference to their implications for health, physical education, and recreation. The companion studies of Bair (2) in physical education and Downey (9) in health education are pioneering attempts to do this. They attempted to identify the basic philosophical positions or assumptions of men and women identified as leaders in their respective fields by asking them to respond to multiple-answer questionnaires containing statements relevant to the philosophical assumptions of idealism, realism, pragmatism, and aritomism[2] and relevant to the logical implication of these assumptions for practices in physical education or health education. Basing their reasoning on the assumptions that (a) philosophical beliefs determine personal actions, and (b) that practices advocated by leaders influence the development of a professional field, Bair concluded: "On the basis of the present indicated beliefs, most professional leaders appear to be providing a predominantly naturalistic direction to American physical education. . . . some evidences of strong spiritualistic beliefs . . . suggested a dual influence and lack of general agreement in some areas of physical education" (2:161). Downey's conclusions concerning health education were substantially in accord with Bair's.

The Validation of Hypotheses. As has been noted above, every sound research study involves the validation of one or more hypotheses. The studies discussed here were selected because the use

[2] Aritomism is a combination of Aristotelian and Scholastic philosophy.

of the philosophical methods is specifically identified in the report of the study.

In the field of child growth, Meredith (17) used the processes of analysis and synthesis with telling effect to establish generalizations about the extent to which secular, socio-economic, ethnic, and regional factors affect the growth of children. His data were hundreds of studies, drawn from a time span of 50 years, which reported the height and weight of children of various socio-economic levels, many different ethnic groups, and several geographic areas. Sorting, selecting, classifying, and comparing the studies for each age-sex group, he succeeded in isolating groups of studies in which three of the factors were "held constant" because they were essentially the same in all of the studies. This enabled him to isolate the effect which might be considered attributable to the fourth factor. In this way he was able to make statements about the probable maximum effect of each factor considered separately.

A study of human strength by Hunsicker and Greey (14) illustrates a different approach to hypothesis testing. They identified a series of questions about strength which are frequently raised by investigators working in that field. They then assembled many studies related to human strength reported by research workers concerned with some specific aspect of the phenomenon of strength. Classifying the findings from these studies under the headings suggested by the questions, Hunsicker and Greey attempted to state the "best" answer to the question in the light of presently available research evidence.

Burke (3) used a similar design to test current hypotheses about the physiological mechanisms which account for the phenomenon called warm-up. He supplemented the findings from the literature with evidence elicited from an experimental study designed to provide relevant information on these points. An attempt to establish valid principles covering the use of progressive resistance exercises as a therapeutic modality, presented by Rasch and Freeman (21), illustrates the use of a similar approach in the field of physical therapy. In physiology, Henry (11) attempted to resolve certain controversies about conflicting values claimed for four experimentally developed tests of cardio-respiratory function with a similarly designed study.

In another study Henry (12) demonstrated the tenability of hypothetical equations for predicting world records in running events, by bringing together knowledge drawn from exercise physiology, competitive sports, and mathematics. He validated his predictive procedure by comparing hypothetical records with actual records in those events which have been extensively and intensively practiced for competitive purposes. The validation of his predictions for the less popular events will be a challenge to future competitors.

Ulrich's study (24) of stress in college women under conditions of competition provides a clear-cut example of the effective use of hypotheses stated in null form to provide a framework within which the significance of many complex relationships may be systematically examined.

An unusually comprehensive and well-defined example of the process of hypothesis testing is Scott's study of kinesthesis (22). On the basis of many speculations about the nature of kinesthesis reported in the literature, Scott formulated six hypotheses, each of which identified some particular aspect of the phenomenon under consideration. Using objective data from many tests involving some element of kinesthetic perception, she analyzed the relationships among these data statistically. She then compared the findings from these analyses with the theoretical findings deduced from the hypotheses. For example, from the hypothesis "muscular contraction of a known amount is a function of kinesthesis," she deduced that each subject should be equally accurate in performing "skills" involving the use of various muscle groups. Analysis of the test results showed that "the use of the arms in adapting effort is apparently unrelated to similar functions in the legs." Since this finding was not in accord with the theoretical findings deduced from the hypothesis, Scott concluded that the hypothesis was not tenable in the light of the experimental evidence. Five hypotheses were similarly rejected. The hypothesis that "learning of a new skill is facilitated by kinesthetic cues which make for similarity of achievement in tasks to be learned" was judged "extremely plausible" because hypothetically related items deduced from it were substantially in accord with similar items in the results of the statistical analysis of the test data.

Studies Emphasizing the Concept of Meaning. These studies differ from those cited above primarily because the hypotheses being investigated are hypotheses about the *meaning* of observed evidence in the lives of human beings rather than about the observed evidence as such. Such studies impose a double burden of interpretation on the investigator; the observed *facts* must first be interpreted in terms of the basic assumptions, and then this interpretation of the facts must be reinterpreted in relation to the investigator's basic assumptions about "what it means to be a human being who finds life meaningful" or the nature of *meaning*, as such.

The meaning of play as "a significant function of living" was explored by Huizinga. His book *Homo Ludens* (Man the Player) was published in German in 1944 and has only recently become available in English translation (13). Beginning with the basic assumption that a persistent, universal function of living such as play always "means something," he has attempted to define that meaning by examining the characteristics of play as they are manifested in the various cultures of the world.

A similar attempt directed toward discovering the meaning inherent in human movement and kinesthesia is currently being made by Ellfeldt and Metheny (10). Accepting the basic assumptions of the contemporary philosophy of symbolic transformation about the difference between the animal "brain" and the human "mind" and the processes through which the human "mind" transforms percepts into concepts, they have analyzed observations of many types of movement with a view to developing and substantiating a general theory of the meaning of movement-kinesthesia as a human experience. Their first paper proposed a tentative hypothesis stated in a vocabulary which was developed from their initial analysis of the elements common to all movement forms. The logical construction of such a vocabulary illustrates a specialized use of philosophical methods.

Studies Emphasizing the Concept of Value. The problem of values is implicit in every attempt to find a logical answer to a controversial or unsettled question. The studies cited here differ from those mentioned in the preceding sections because the hypotheses being investigated are hypotheses about the *value* implicit in or attached to observed facts which have been interpreted as *facts* and reinterpreted in terms of their *meanings* in the context of certain basic assumptions about the nature of meaning in human lives.

Cobb (5) stated four possible hypotheses about the *values* inherent in various physical education situations in an attempt to determine "the framework within which college physical education functions *ought* to operate" (4:144). She "stated the question, explained its controversial nature, the historical background out of which it arose, and reported various solutions that had been suggested by presumably responsible and qualified persons" (4:143). She also sought in the areas of physiology, psychology, sociology, and education for information which might provide some basis for making decisions among the disparate solutions. Thus, she determined which one of the four hypotheses about *value* was most tenable in the light of all of the available evidence as interpreted in terms of *meanings* associated with it by experienced personnel in several fields concerned with various aspects of human lives.

The "Statement of Policies and Procedures for Competition in Girls and Women's Sports" (1) is an example of the outcomes of a study of values carried on by a large group of people over a long period of time. They were attempting to identify, i.e., find the "best answer," to the questions about "what is 'right' for girls in one important area of their lives" by resolving "the confusing and often contradictory issues concerning values of athletic competition for girls in our present social order" (18). A similar example of an attempt to identify values by co-ordinating the thinking of many people is the report of the Educational Policies Commission on *School Athletics* (19).

BIBLIOGRAPHY

1. American Association for Health, Physical Education and Recreation, Division for Girls and Women's Sports. "Statement of Policies and Procedures for Competition in Girls and Women's Sports." *Journal of Health-Physical Education-Recreation* 28:6:57-58; September 1957.
2. Bair, Donn. *Identification of Some Philsophical Beliefs Held by Influential Professional Leaders in American Physical Education.* Unpublished Doctor's thesis. Los Angeles: University of Southern California, 1956. Also, see *Research Quarterly* 28:315-20; December 1957.
3. Burke, Roger K. *Relationships Between Physical Performance and Warm-Up Procedures of Varying Intensity and Duration.* Unpublished Doctor's thesis. Los Angeles: University of Southern California, 1957. Also, see "Effect of Motivational Stress upon Physical Performance" by Celeste Ulrich and Roger K. Burke in the *Research Quarterly* 28:403-12; December 1957.
4. Cobb, Louise S. "Philosophical Research Methods." *Research Methods Applied to Health, Physical Education, and Recreation.* Washington, D. C.: American Association for Health, Physical Education, and Recreation, a department of the National Education Association, 1949.

5. COBB, LOUISE S. *A Study of the Functions of Physical Education in Higher Education*. New York: Bureau of Publications, Teachers College, Columbia University, 1943.
6. COMMINS, SAXE, and LINSCOTT, ROBERT N., EDITORS. *The World's Great Thinkers. Man and Man: The Social Philosophers*. New York: Random House, 1947.
7. COMMINS, SAXE, and LINSCOTT, ROBERT N., EDITORS. *The World's Great Thinkers. Man and Spirit: The Speculative Philosophers*. New York: Random House, 1947.
8. COMMINS, SAXE, and LINSCOTT, ROBERT N., EDITORS. *The World's Great Thinkers. Man and the Universe: The Philosophers of Science*. New York: Random House, 1947.
9. DOWNEY, ROBERT JAMES. *An Identification of the Philosophical Beliefs of Educators in the Field of Health Education*. Unpublished Doctor's thesis. Los Angeles: University of Southern California, 1956.
10. ELLFELDT, LOIS E., and METHENY, ELEANOR. "Movement and Meaning: Development of a General Theory." *Research Quarterly* 29:264-73, October 1958.
11. HENRY, FRANKLIN. "Functional Tests III: Some Effects of the Common Cold on Cardio-respiratory Adjustments to Exercise and Cardio-respiratory Test Scores." *Research Quarterly* 13:185-200; May 1942.
12. HENRY, FRANKLIN. "Prediction of World Records in Running Sixty Yards to Twenty-Six Miles." *Research Quarterly* 26:147-58; May 1955.
13. HUIZINGA, JOHAN. *Homo Ludens, A Study of the Play Element in Culture*. Boston: Beacon Press, 1955.
14. HUNSICKER, PAUL, and GREEY, GEORGE. "Studies in Human Strength." *Research Quarterly* 28:109-22; May 1957.
15. JASPERS, KARL. *The Perennial Scope of Philosophy*. New York: Philosophical Library, 1949.
16. LANGER, SUSANNE K. *Philosophy in a New Key*. New York: New American Library, a Mentor Book, 1948.
17. MEREDITH, HOWARD V. "Stature and Weight of Children in the United States." *American Journal of Diseases of Children* 62:909-32; 1941.
18. METHENY, ELEANOR. "Relative Values in Athletics for Girls." *Journal of Educational Sociology* 28:268-70; 1955.
19. NATIONAL EDUCATION ASSOCIATION and AMERICAN ASSOCIATION OF SCHOOL ADMINISTRATORS, EDUCATIONAL POLICIES COMMISSION. *School Athletics, Problems and Policies*. Washington, D. C.: the Commission, 1954. 116 p.
20. OPPENHEIMER, J. ROBERT. *Science and the Common Understanding*. New York: Simon and Schuster, 1954.
21. RASCH, PHILIP J., and FREEMAN, RICHARD. "The Physiology of Progressive Resistance Exercise: a Review." *Journal of American Association for Physical and Mental Rehabilitation* 8:35-41; March 1954.
22. SCOTT, M. GLADYS. "Measurement of Kinesthesis." *Research Quarterly* 26:324-41; October 1955.
23. SELYE, HANS. *The Stress of Life*. New York: McGraw-Hill Book Co., 1956.
24. ULRICH, CELESTE. "Measurement of Stress Evidenced by College Women in Situations Involving Competition." *Research Quarterly* 28:160-72; May 1957.
25. WHITE, LYNN, EDITOR. *Frontiers of Knowledge in the Study of Man*. New York: Harper & Brothers, 1956.
26. WHITEHEAD, ALBERT NORTH. *Science and the Modern World*. New York: New American Library, a Mentor Book, 1948.

CHAPTER 16

Writing the Research Report

CAROLYN BOOKWALTER
KARL BOOKWALTER, MARJORIE PHILLIPS

THE FUNCTION OF THE WRITTEN REPORT IS TO COMMUNICATE A set of ideas or facts to the reader and, if the report is to be effective, the ideas must be conveyed in a form that is clear and easily understandable.

Carelessly written, poorly organized, or faulty reports will so distract the reader's attention that he will find it difficult, if not impossible, to absorb the content. Any writer will enhance his professional contribution when he acquires the skills requisite to effective reporting.

CHARACTERISTICS OF A GOOD REPORT

Unity, coherence, and emphasis are involved in the proper organization of a report. Clarity, correct presentation of material, completeness, and conciseness are other characteristics of the good research report.

To attain unity, coherence, and emphasis, each of the major sections should appear in logical order. One section should lead naturally to the next. The development of ideas within sections must be systematic, with care being taken that no omissions appear in the reasoning or trend of thought. Proper emphases should be given to important topics through the arrangement of materials and through expression of important ideas so that the reader is immediately aware of them.

To obtain clarity in the report, the writer needs to have a comprehensive understanding of all factors related to his problem. He cannot hope to guide the reader to a successful and full understanding of the problem if his own thinking is confused and uncritical. All statements should be expressed in terms that eliminate any doubt from the mind of the reader as to the exact meaning intended.

Correctness of presentation is dependent upon a knowledge of diction, rhetoric, grammar, spelling, punctuation, bibliographical and footnote form, form of tables and graphs, and spacing of materials. An otherwise valuable piece of research may be badly damaged in the reporting with such practices as careless vocabulary usage, violation of the rules of grammar and spelling, and lack of unity and coherence in sentence structure.

The report must be comprehensive; all the facts of the investigation should be quite clear. The novice writer frequently forgets that many essential facts which are familiar to him may not be at all obvious to the reader. The reader should not be left with unanswered questions. This does not imply that the report must necessarily be lengthy, since it must also meet the criterion of conciseness. All irrelevant ideas and superfluous material should be eliminated, so that the report is as brief as is consistent with completeness and clarity.

PRELIMINARY STEPS

Many writers have found that the best results can be obtained by preparing an outline preliminary to writing. If the research study has been well planned from the reading of related literature through the interpretation of data, the writing of the report will not be difficult. Some research workers have found it desirable to plan the writing of the report at the time the plan is developed for the research study. Such research workers find that they are able to organize, process, and analyze their data so that the materials may be written up with a minimum of time and effort.

The outline of the major topics is followed by subdivisions under each topic. The brief is much more detailed than the outline and contains all the items to be discussed under each topic. The brief grows out of the preliminary outline and represents a far more advanced stage of the writer's planning.

Considerable time should be given to the preparation of the brief to ensure logical arrangement of topics, proper emphasis of materials, and inclusion of everything important in the report. The writer needs to determine where the materials will be placed and how they will be presented.

From the brief, the preliminary draft of the report is written. This draft, in turn, should be carefully revised and corrected before the final copy is prepared. It should be examined for clarity, arrangement of material, correctness of presentation, completeness, and conciseness. It is usually wise to ask one or two qualified individuals to read the report and suggest improvements before it is finally submitted for publication.

MAJOR DIVISIONS OF THE REPORT

There is no one form of arrangement which will meet the needs of every situation. The sections of the report should be organized according to the nature of the problem and the purpose for which the report is written.

Since practices vary among the periodicals, publishing houses, and educational institutions, it is essential that the writer acquaint himself with any specific policies required by the institution or journal for which he is writing.

Materials may be arranged by parts, chapters, center heads, and side heads. The arrangement selected depends upon the length of the report and the type of research study. Short papers (less than 50 pages) are frequently written with only center and side heads.

Extended research reports commonly include the following parts:

1. *Preliminary material*—title page, preface, table of contents, list of tables, and list of figures.
2. *Body of the report*—formulation and definition of the problem, delimitations, limitations, related literature, sources of data, procedures, presentation and interpretation of findings, summary, conclusions, and recommendations (when appropriate).
3. *Supplementary material*—bibliography, appendix, and sometimes an index.

Preliminary Material. This consists of the following:

Title Page. The title page includes the title of the report, the author's name, his institutional connection, the place, and the date of writing. All material should be well arranged and centered between the margins of the page. No terminal punctuation is used on the title page.

The title should clearly indicate the subject, be reasonably brief, and be correctly worded. Good (11:179) suggests that certain forms of expression should be avoided. Among those mentioned are: "Some Aspects of . . . "; "A Study of . . . "; "A Scientific Study of . . . "; and "An Analysis of . . . ." An example of a title is "The History of Physical Education for Women in the United States, 1700-1950." Such a title gives the subject, the research method, the place, and the time.

Inclusion of the date of the writing of the research report on the title page is important, since the reader may then evaluate the results of the investigation in light of previous or more recent findings.

Preface. The preface may or may not appear in the report. It is sometimes used to express the writer's personal interest in the problem, or it may contain acknowledgments to people to whom the writer is indebted. It is advisable for acknowledgments to be made only to those who have rendered considerable assistance of nonroutine nature, and it is desirable at all times for the wording to be brief, simple, and impersonal.

Table of Contents, List of Tables, and List of Figures. These parts of the report are invaluable as mechanical aids to the reader, not only in acquainting him with the study but in enabling him to locate material. The listing of topics in the table of contents should be sufficiently detailed to allow the reader to find readily any section in which he is interested. The relationships between topics and subtopics can be emphasized by enumeration, indentation, and alignment. The lists of tables and figures should give the number of the table or figure, the exact and complete title, and the page number on which the table or figure may be found.

The typing form varies in writing manuals, but there is agreement that the titles "table of contents," "list of tables," and "list of figures" should be consistent in spacing vertically and horizontally to separate major and minor parts. In some manuals, the form for capitalization of titles changes when the titles are listed in the list of tables or the list of figures.

Chapters or Section Headings. The introductory chapter or section should lead the reader to the problem. It may include the statement of the problem, definitions of words which might have several meanings to the reader, limitations of the problem, delimitations of the problem, need for the study, nature of the data, presentation of basic assumptions, a short description of the procedure for the study, and related literature. If the material for the related literature is rather extensive, the related literature may be placed in a chapter by itself. The description of the procedure may also be placed in a chapter by itself if the material seems to warrant the space.

It is well to state the problem in the opening paragraph of the introduction. This serves to fix the reader's attention and facilitate his understanding of subsequent discussion. The problem must be stated clearly and logically. The statement of the problem should not be worded as a purpose nor as an action. "To make," "to determine," and other acts should not be a part of the statement. It should be written in sentence form. In formulating the statement of the problem, the writer may bring out the weakness of a declarative sentence by putting the statement first in question form.

The title and the statement of the problem are not necessarily alike. Following the statement of the problem, there may be a need for a brief explanation or interpretation. Any specific questions to be answered should be carefully outlined.

Only words which could be misinterpreted should be defined. For instance, "junior high school" could mean 7th and 8th grades, 8th and 9th grades, or 7th, 8th, and 9th grades. It would be necessary for the writer to define what he means by "junior high school." In some cases, technical terms should be defined.

In *delimitations*, the author includes information about what the study covers as to scope, sources of data, and methods. Sometimes the phrase, "definition of the problem," is used in place of the term "delimitations."

In the *limitations*, the author should state the weaknesses of the study as to scope, sources of data, and interpretations. Any uncontrollable circumstances which may have affected the results should be explained and their possible implications pointed out. Methods or data which were found to be valueless should be mentioned, since this information may be of great assistance to an investigator planning a similar study.

Findings. The findings of the study may be reported in one or more chapters. The nature of the study helps the writer determine how many chapters are needed to present the findings in a clear and accurate manner. In the material on findings, the author must decide to what extent the original data should be displayed, how data will be organized for presentation, how data will be displayed, what *format* should be used for the tables and for the figures. Some material is more appropriately placed in the appendix than in the findings.

Analysis and Interpretation of Data. The analysis and interpretation of data are facilitated through the use of tables, graphs, statistical computations, and discussion. Through the discussion, the data are presented more analytically and with more emphasis on facts of importance than can be accomplished through tables and graphs. This does not involve a simple enumeration or recapitulation of what may have already been illustrated by graphical methods. The discussion must be developed systematically in the chapters on findings, so that each chapter leads logically to the next and so that the import of the interpretation leads naturally to the conclusions. Agreements, differences, and relationships should be shown and discussed.

The discussion in a typical paper should concern itself with the source and magnitude of errors, by-products of the research, suggestions for improvement of procedure, the applications of the findings, and the true meaning of what has been found. It is essential that the investigator have a complete understanding of his data and what they mean, for if he lacks the necessary insight, his interpretations can neither be accurate nor properly revealing. If the results are in disagreement with similar research by other workers in the field, the discrepancies should be pointed out and possible explanations given. In no case should any hint of bias appear, and all facts should be given in a straightforward exposition of what has been found.

Summary, Conclusions, and Recommendations. In extensive reports, a summary placed at the end of each chapter is of great assistance to the reader. The summarizing paragraph at the end of each chapter is usually stated in general terms and in terms of the general implications of the data as well as of the objectives which were established for the investigation. It provides the reader with a comprehensive understanding of what has been accomplished in the particular section of the study, and clearly states significant results.

The summarizing paragraph or chapter at the end of the report is usually stated in general terms and gives an overview of the entire study. The summaries of the chapters on findings should be analyzed to find agreements, differences, and interrelationships of the findings. The writer should avoid repeating chapter summaries word for word in the summary chapter at the end of the report.

The conclusions are stated in specific terms and should provide explicit answers to the questions posed in the problem. The conclusions must be substantiated by the data and must be free from unsupported opinion. Any qualifications or limitations should be carefully explained.

The relation of the conclusions to the statement of the problem can best be illustrated by a specific example.

> *Statement of the problem*—Meyer and Pella (18): ". . . the effects of hard laboratory exercise on the total and differential blood count of normal, young, adult women." *Conclusions*: "1. Hard laboratory exercise of short duration produces leucocytosis in normal, young, adult women. 2. Post-exercise leucocytosis is repeatable. 3. Post-exercise leucocytosis may be followed by neutrophilic leucopenia, with a subsequent return to pre-exercise level. 4. There is a highly significant increase in lymphocytes and a significant decrease in neutrophils in the immediate post-exercise differential counts."

Recommendations may follow the conclusions, but a differentiation should be made between recommendations that are based upon data and those that are derived from the judgment of the writer.

Supplementary Material. The supplementary material consists of the bibliography, the appendix, and possibly the index. The bibliography should consist of selected references the author has found useful in the solution of the problem. All sources to which the author refers in the body of the report should be included in the bibliography, except for personal letters or sources not available to the reader through library services. The inquiring reader may examine the bibliography as one means of judging the discrimination of the writer and his scholarly grasp of his subject.

Documentation of each quoted fact and opinion is essential. This provides an acknowledgment of the writer's indebtedness to his sources and provides a point of departure for an investigator wishing to develop the problem further. Plagiarism must be avoided. Changing only one word in a sentence or a paragraph is not a means of avoiding plagiarism. Means of avoiding plagiarism are expressing the idea in one's own words completely, documenting accurately, quoting accurately, and obtaining permission from the publishers or author for the use of the quotation. The latter is particularly necessary if the materials are to be reproduced in any form, such as microcarding, printing, and other methods. Quoted materials, facts, and opinions obtained from sources may be documented by footnotes or footnote reference to the bibliography. When permission is granted by the publisher or author for quotations, the permission must be indicated in the footnote.

There is no standard form for bibliographical material and footnotes which is universally accepted; therefore the writer should conform to the policies and standards demanded by the publisher, editor, or institution for whom he is writing. Whatever form is chosen, the writer must be consistent in following this form throughout his report. An annotated bibliography, while not essential, is very useful and is further evidence of the care with which the study is prepared.

Anderson and Valentine (1:370-71) have devised an extensive list of rules for arranging a bibliography. Their complete list is not duplicated here, but some of the more commonly needed rules are mentioned.

1. Enter an article by a married woman under the name which appears on the article for which the reference is made.
2. Enter names beginning with M', Mc, Mac, or St., and Ste., whether the following letter is capitalized or not, as though the prefix were spelled out in full as Mac, or Saint, as in the following examples: M'Intyre, J., and Mackey, J.
3. Enter compound names under the first part of the name.
4. If an author has more than one publication within a single year, the publications within the year are arranged alphabetically by title, disregarding the articles of speech such as "the" or "an."
5. References to joint articles follow articles by the senior author alone.
6. If the reference is an article or section prepared by a particular author, but published in a collection of articles such as a handbook or an annual review, the entry is made under the particular author.
7. If reference is to an entire book which is a collection of articles by various authors, entry is made under the name of the editor, if it is clear that he is responsible for assembling the material.
8. Reports of committees, governmental organizations, etc., are grouped together at the end of the alphabetical list and are arranged alphabetically by title.

The *appendix* contains all materials which are too unwieldy or cumbersome to include in the text but which will aid in the understanding of the report. All materials included in the appendix must appear in their original form.

The appendix may include such materials as catalogues, courses of study, checklists, evaluation sheets, form letters, questionnaires, lists of co-operating individuals or institutions, computation sheets, raw data, formulas, and documents. Sometimes authors may place tables and figures in the appendix, in which case they should be numbered in the order of their actual appearance. When this is done, reference is made in the text of the report to the tables or figures in the appendix. Usually the less important tables and figures are placed in the appendix. The author should avoid using tables and figures as fillers. Each should have a reason for existence.

TABULAR AND GRAPHIC PRESENTATION

Tables. Tables provide for the orderly arrangement of data in columns and rows according to one or more pertinent classifications of the subject matter. Tables thus set up give more clarity and and meaning to data and make for more ready interpretation. For example, a rank order of states by such factors as population, expenditures, or wealth is far more meaningful than their alphabetical arrangement. Customarily the highest or most desirable value is placed at the top, and lower values succeed in descending order.

Placement. When practicable, tabular material should be placed between logically related paragraphs. Tables should not interrupt a sentence. The discussion preceding a table should introduce the table and the material which follows should interpret the table. Preceding material should be separated from the table by four spaces and subsequent material should follow the table three spaces below. It is desirable to confine most tables to one page or less. If the table is reproduced to conform to the typed or printed page, care must be taken that the size of the original lettering is such that it can be read in the reduced form. If the table is wider than it is long, it may be placed crosswise on the page with the title toward the left of the page. The table designation and title should be at the top of the table. The designation should be in capital letters and numerals, and the title should be capitalized and follow the designation period by two spaces. The title and designation should never exceed the width of the table; if narrower, it should be centered over the table. If there is a carry-over to the title, the carry-over should be typed single space and indented two spaces. Such expressions as "table showing," or "distribution of," should be avoided and abbreviations should not be used in the title if they can be avoided. From the title the reader should know what data are involved and where and when they were obtained, and the table should be self-explanatory.

A Simple Table. From the example below and those to follow, it can be seen that there are various styles for making up tables and figures. Throughout this chapter, these inconsistencies are permitted because the material is being reproduced exactly. The examples were chosen for their general merit and for their exemplification of certain basic principles. They were consistent in their original sources. Consistency with an acceptable style (usually that of the publication in which the author expects his report to appear) is more important than what the style is which is accepted.

TABLE 1.—MEAN TIMES AND STANDARD DEVIATIONS IN SECONDS FOR ALL 220-YARD TRIALS

Condition	Mean	Standard deviation
Experimental	24.33	2.05
Control	24.26	2.14
Free	24.28	2.06

FIGURE I. A simple table (Source: *Research Quarterly* 30:134; May 1959.)

Boxheads are column captions which tersely define the data below. The stub, which is the column of row titles below the boxhead, is left open on both sides of the table. Brace headings are captions which relate to two or more boxheads and should be centered and limited by the extension of the outer vertical lines of the columns involved. When footnotes are needed to clarify exceptions, they should follow the lower line of the table by a double space, with a single space between footnotes.

Analytical Table. In Figure II, the title, "Distribution of Scores and Reliability Coefficients for the Study Groups on the Instrument to Measure Locomotor Response (Rhythmeter)," seems to be in violation of the rule to avoid the use of such introductory phrases as "distribution of." Here, the manner of distribution is of import and the title is appropriate.

TABLE 1.—DISTRIBUTION OF SCORES AND RELIABILITY COEFFICIENTS FOR THE STUDY GROUPS ON THE INSTRUMENT TO MEASURE LOCOMOTOR RESPONSE (RHYTHMETER)

Group	Number of cases (169)	Mean	Range (30) items	S. D.	Hoyt's formula (r)	Split-half correlation
Control Group						
General college population	89	8.1	1-27	6.5	.91	.89
Experimental Group						
Dance Club members	42	20.3	13-28	4.5	.68	.89
Professional dancers	383	21.7	12-29	4.4	.575	.87
Combined Experimental Group	80	20.9	12-29	4.4	.66	.88

FIGURE II. An analytical table (Source: *Research Quarterly* 29:345; October 1958.)

In the table, the groups are alphabetically arranged. The sub-groups under "Experimental Group" are alphabetically arranged except for the combined group. A miscellaneous group would be listed last in other tables. It may be possible in some tables to order sub-groups by magnitude of scores with the highest scores at the top. In other tables, it may be desirable to give totals for the columns and the rows. It is essential that the sideheads in the stub of the table be co-ordinated in value.

As a rule, not over three or four classifications should be involved in one table, as meaning is lost with complexity.

It should be noted that the title is fully capitalized and that the carry-over has been centered to follow the style of the *Quarterly*. Note that regular lines begin and close this table. The stubs for each of the first and last stories of boxheads are open in this table. This is preferred to closing the sides of the table.

Graphs. Graphical methods facilitate the drawing of logical conclusions and inferences.

The Co-ordinate System. Basically, a graph is obtained by plotting figures in relation to a vertical (Y) axis or ordinate and a horizontal (X) axis or abscissa, the intersection of which forms the origin or zero of both axes.

Scores or values to the right of the origin are positive, those to the left are negative. Frequencies or values above the origin are positive. Those below are negative. Thus, the signs of all products in quadrants I and III are positive, while those in quadrants II and IV are negative. Most ordinary graphs and figures using the co-ordinate system employ only quadrant I. Captions should accompany both the vertical and the horizontal scales. A graph should be planned so that it is attractive, simple, self-explanatory, and accurately informative.

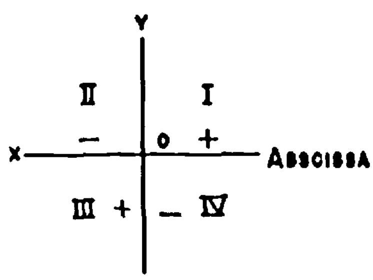


FIGURE III. The axes and quadrants upon which the graphs are based.

Usually, a figure is reduced from the original drawing for printing. Care must be taken that figures, lettering, and especially numbers, are large enough to be readable when reduced. If a scale accompanies the drawing, the scale should be in absolute proportion rather than in inches, centimeters, etc.

Frequency Polygon. In this figure the frequencies or dependent variables are usually placed on the Y axis or ordinate, whereas the independent variables or score values are placed on the X axis or abscissa. The height of the figure should be in ratio of about 3 to 4 with the width. Such ratios may vary from 4 to 5 (.80) to 4 to 7 (.57), however. If several figures are to be compared, they should have a common ratio.

The ranges of the scores and of the frequencies are determined and convenient intervals are laid off on the respective axes. Both axes should extend at least one interval beyond each limit of the data. Captions should accompany both scales.

Dots are first placed midway of the score intervals at the height of the appropriate frequencies shown on the ordinate. These dots are then connected by straight lines, and the end dots are connected to the base line by lines drawn to the midpoints of the nearest interval of zero frequency.

When two or more distributions are to be compared on the same figure, dotted or broken lines should be used to differentiate. The frequencies should be reduced to percentages, so that unequal populations will not distort the comparisons. When only one distribution is shown, the frequencies should be shown on the scale on the left and the percentages may be shown in the vertical scale on the right. It is undesirable to draw separate figures for frequencies and percentages.

The exemplary frequency curve in Figure IV was originally plotted on semi-logarithmic paper. If differentiated vertical lines had been drawn at the mean for each polygon, the true increase could be observed.

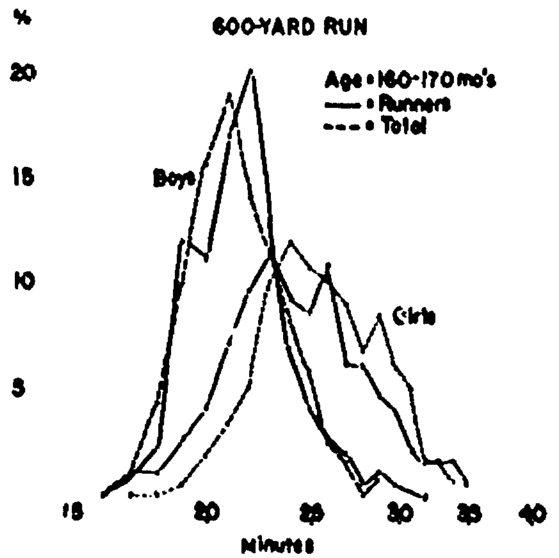


FIGURE IV. Frequency polygon (Source: *Research Quarterly* 29:345; March 1951.)

Histogram. The purpose of a histogram is similar to that of a frequency polygon, to indicate the distribution of the sample with regard to a variable. The histogram differs from the frequency polygon essentially in one respect. This difference is that instead of assuming the midpoint of an interval to be most representative of the cases therein, the cases are assumed to be evenly spread over the interval. Accordingly, at the ordinate height corresponding to the frequencies of a given score or value interval, a horizontal line is drawn the width of the interval parallel to the abscissa. These adjacent horizontal lines are connected by vertical lines. Thus a polygon is outlined, the area of which is proportional to the population of the sample.

Figure Va illustrates an outline histogram in which the percentages of the distribution population are indicated rather than the frequencies, and the mean for the entire college population is shown by the vertical line. Close comparisons of interval fre-

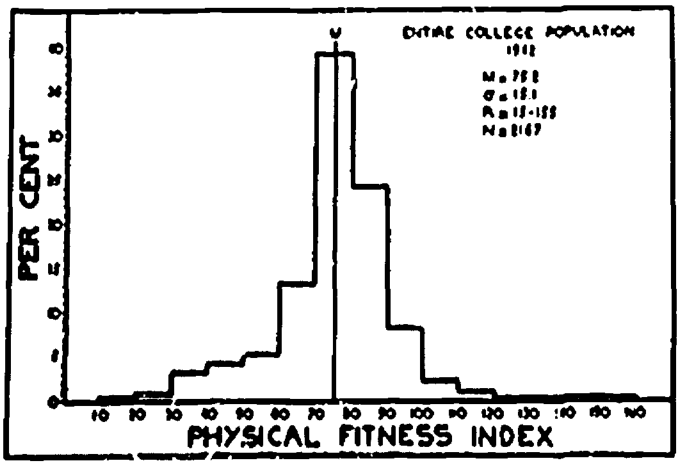


FIGURE Va. Outline histogram (Source: *Research Quarterly* 15:214; October 1944.)

quencies are not readily discernible. The bar histogram corrects such a deficiency, as can be seen in Figure Vb. The bars in the graph are directly proportional to the interval frequencies. The difference between these histograms is made by continuing the connecting vertical lines down to the base line or abscissa. In the illustrative graph, the number of men (frequencies) is indicated on the vertical scale on the left. It should be noted that in each scale the range is correctly greater than the highest expectancy.

Line Graphs. The line graph is usually employed to disclose trends in continuous data. The typical learning curve is an example in which the number of trials are indicated on the abscissa scale and the responses, successes, or scores are indicated on the ordinate scale. The dots for any point of reference are determined by moving out on the abscissa the required distance (number of trials, or the time in years, months, or days) and above that place

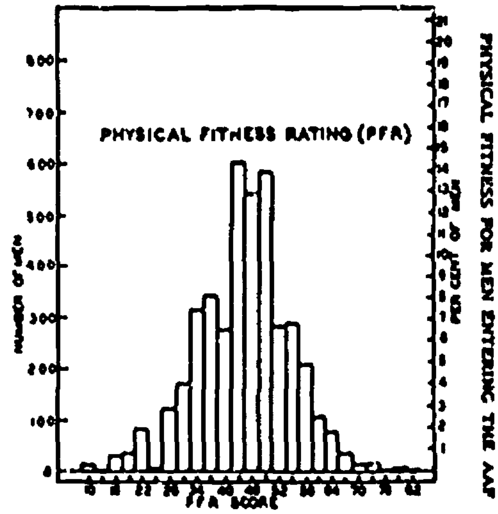


FIGURE Vb. Bar histogram (Source: *Research Quarterly* 17:189; October 1946.)

move to the appropriate height on the ordinate (successes, responses, or score). These dots are connected serially.

In Figure VI, data are presented regarding the learning of three groups (X, Y, and Z). The learning curves are differentiated by

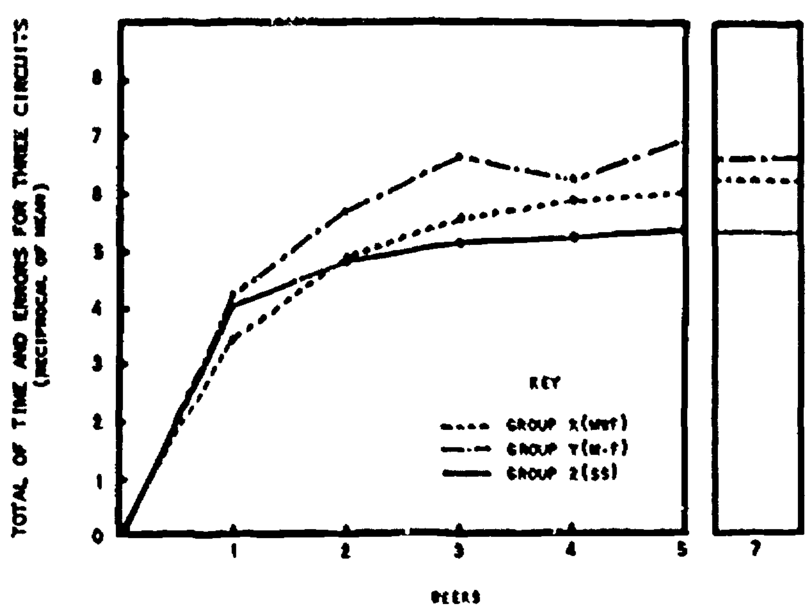


FIGURE VI. Line graph (Source: *Research Quarterly* 30:195; May 1959.)

dotted lines, solid lines, and dot-dash lines respectively. The typical rise followed by a plateau is observable.

Bar Graph. Bar graphs may be vertical or horizontal as in Figure VII. In the horizontal bar graphs shown here, there are several splendid features. The bars are ranked in descending order for area scores—the states being taken out of alphabetical order. The mean for the 25 states is shown by a vertical line. The space between bars is properly about one-half the width of the bar for contrast in reproduction.

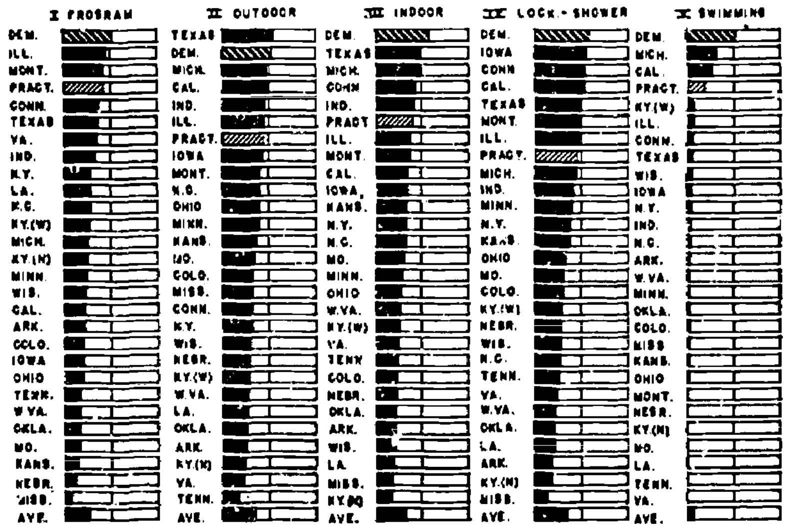


FIGURE VII. Horizontal bar graph (Source: *Professional Contributions No. 4*, American Academy of Physical Education, 1955, p. 6.)

In the preparation of such graphs for reproduction, the worker is warned not to have a dull typing process and a brilliant black India ink value on the bars. Such contrasts are difficult to photograph evenly. The field upon which the bars are drawn should have a minimum of co-ordinate lines, as the latter tend to obscure the picture. This is even more true when bars are in outline rather than in solid form. Color filters may be affixed to cameras to screen out extraneous lines on most co-ordinate paper. The use of blue lined paper will obviate this problem.

No figures of percents or frequency should be written at the ends of the bars. This practice distorts the visual lengths. Instead, the numbers may be written within the bars or indicated by scales at the sides.

Divided Bar Graphs. The general principles for bar graphs apply in this case as well. In addition, the parts of the bar should be in the same order in all bars for easy comparison. If possible, that order should be the one of importance or size. Bar graphs lend themselves well to item comparisons of expenditures, proportional frequency of practice, and sources of total income. They may be substituted generally whenever more than one variable is contrasted for geographical or temporal considerations. Legends should usually accompany the graph to indicate the meaning of bar parts. Such a legend may well be inserted in an open space on the graph. The cross hatching and shading for bar graphs may be purchased in sheets, cut, and pasted on the original graph as needed.

The divided bar graph (with central tendency and variability) is a device for showing the central tendency (mean or median) for such values as rating, item cost, and age at which degrees are attained. It conveniently indicates a representative result, a total

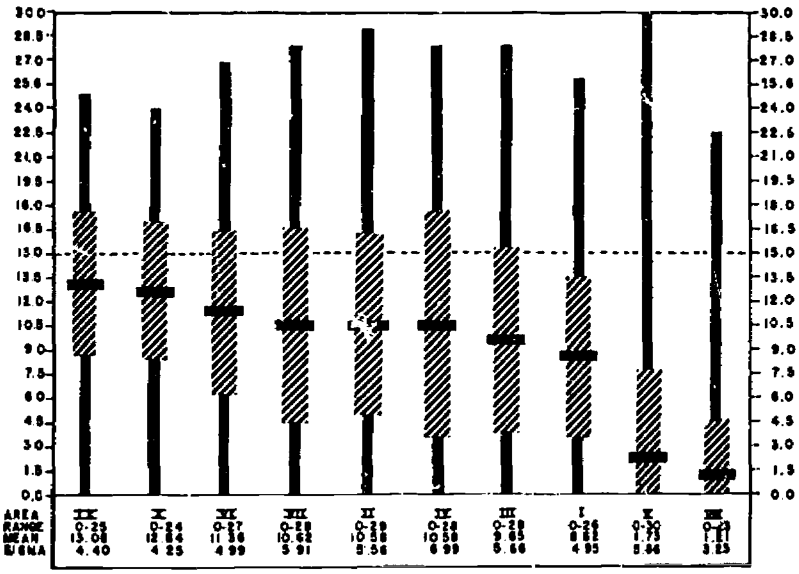


FIGURE VIII. Divided bar graph (Source: K. W. Bookwalter, *National Survey of Health and Physical Education Programs for Boys in High Schools, 1950-54.* Indiana University, 1958, mimeographed.)

range, and a more reliable measure of spread around the central measure, usually sigma or Q respectively.

In Figure IX, the central tendency employed is the mean and the bars are used to indicate the directions of the variabilities from the mean bars. An additional desirable feature is the drafting of the items on the left in black India ink to avoid tone contrast with the bars.

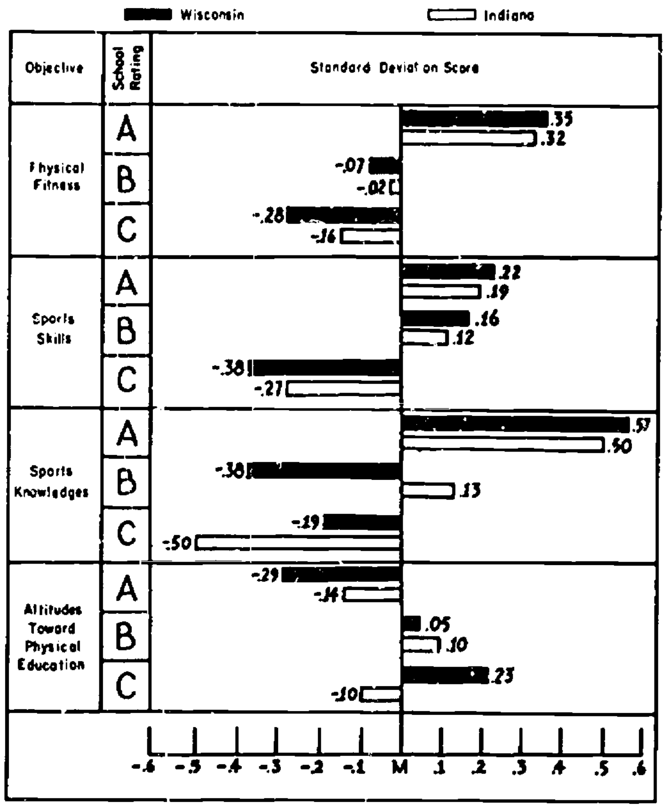


FIGURE IX. Divided bar graph. (Source: Doctor's thesis, Indiana University by James Joseph Rice, entitled *Status in Health and Physical Education Score Card Number II Standards Compared with Selected Outcomes in Physical Education*, 1957, p. 55.)

Circle Graph. For depicting percentage and source of certain practices, item distribution of costs, or expenditures, the circle graph is a valuable technique.

In the illustration certain basic criteria are generally followed; namely, a vertical line starts the comparisons which move clockwise, the items are arranged in descending order, the items are identified, and all items in the two circles are compared in the same order. This last criterion is more important than the criterion that the items should be in rank order, in case several circles are being compared.

In determining the sizes of the sectors, measurements should be in degrees and be proportionate to the 360 degrees of the circle. A protractor is used in such measurement.

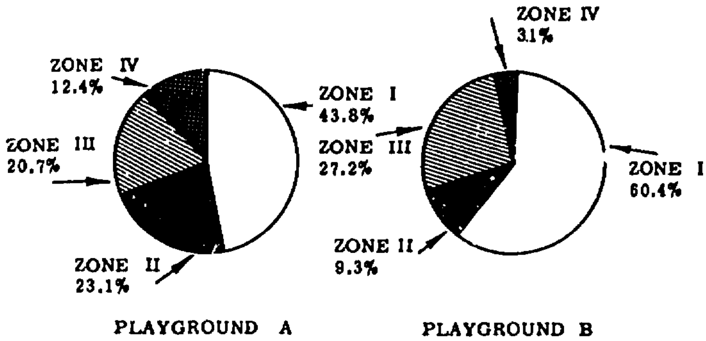


FIGURE X. Circle graph (Adapted from Chart II. Patron Participation by Zones, *Research Quarterly* 28:144; May 1957.)

Pictograph. The pictograph technique employs stylized drawings, silhouettes, or representations of the objects or values being compared. In function, pictographs are frequently used instead of simple bar graphs. However, they are also used to characterize the nature of certain occupational, age, or commodity groups in geographical or other groupings. Their chief advantage lies in their interest-arousing nature. Certain characteristics of the sample or group may be more vividly depicted by these figurettes also.

The elements utilized in this model and many similar elements can be purchased from art supply houses or bookstores in sheets and then cut out and pasted on the original graph as needed. The legend here properly indicates the unit or units represented by each figure to facilitate interpretation of the chart. The specific values can rarely be obtained from such charts but the general impression is graphically shown and, in any case, confirming tables should accompany the chart.

Pictographs have been nicely combined at times with the use of maps to illustrate distributions of such variables as vocations. products, attitudes, or expenditures.

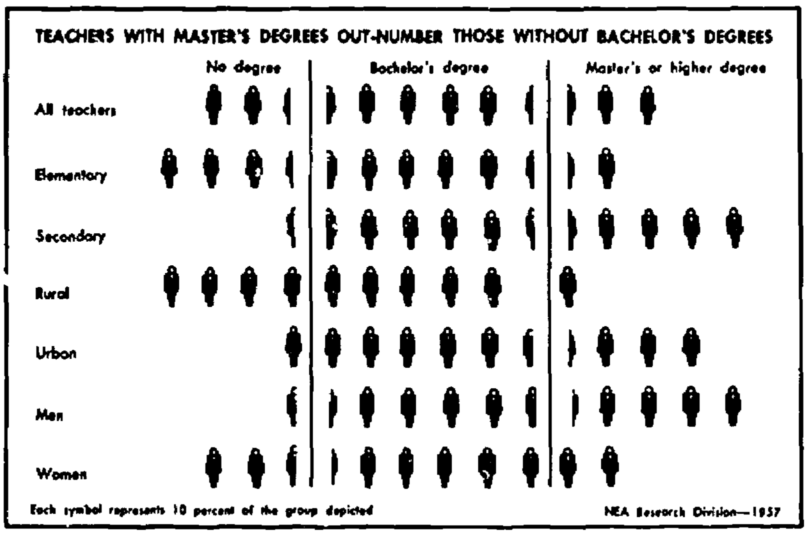


FIGURE XI. Pictograph (Source: *NEA Research Bulletin* 35:1:13; February 1957.)

Diagrams or Pictures. Drawings of objects, testing or laboratory procedures, or mechanical set-ups are examples of uses for the diagram. As a rule, diagrams are substitutes for pictures. There is no accepted rule as to when to use a diagram and when to use a picture. The diagram may be merely a careful sketch, a more exact artists's drawing, or a draftsman's sketch that follows engineering requirements. The nature of the report and the type of reader expected will determine the exactitude. Sometimes the drawing may be more satisfactory than the picture, and at other times the reverse is true.

In the case of Figure XIIa, the tensiometer is an object rather well depicted by the careful artists's sketch. Only good photography could have assured such detail. On the other hand, Figure XIIb is an example of a diagram of a test set-up which failed to give the detail of the light over the mirror, the shield before the mirror, the electric counter, the dry battery, the stylus, and the

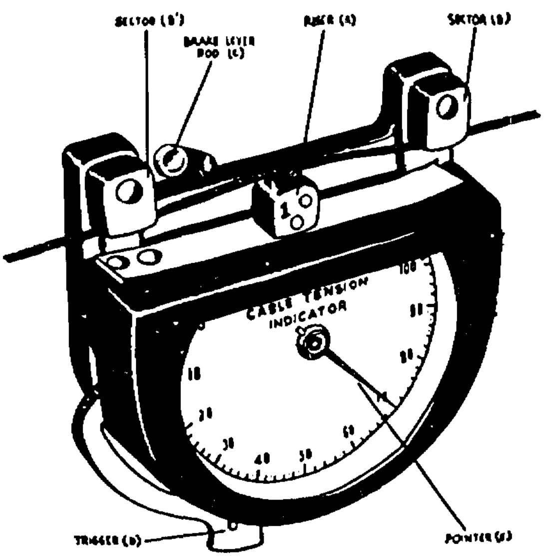


Figure XIIa. Sketch picture of the tensiometer (Source: *Research Quarterly* 19:121; May 1948.)

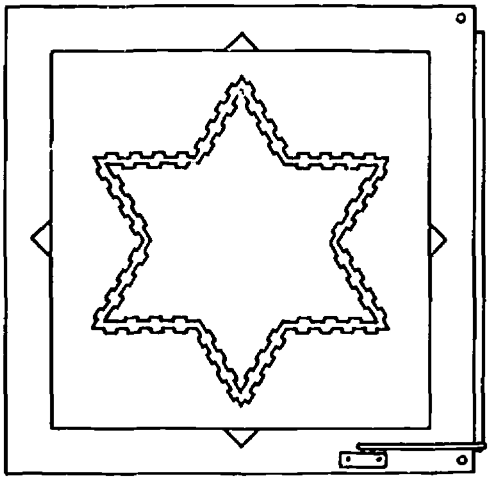


Figure XIIb. Diagram of mirror-drawing instrument (Source: *Research Quarterly* 30: 192; May 1959.)

timer—all parts of the set-up for the stabilimeter. The diagram was later replaced by the commercial photograph which is shown in Figure XIIc. Note how a black background brings out the highlights of the experimental set-up.

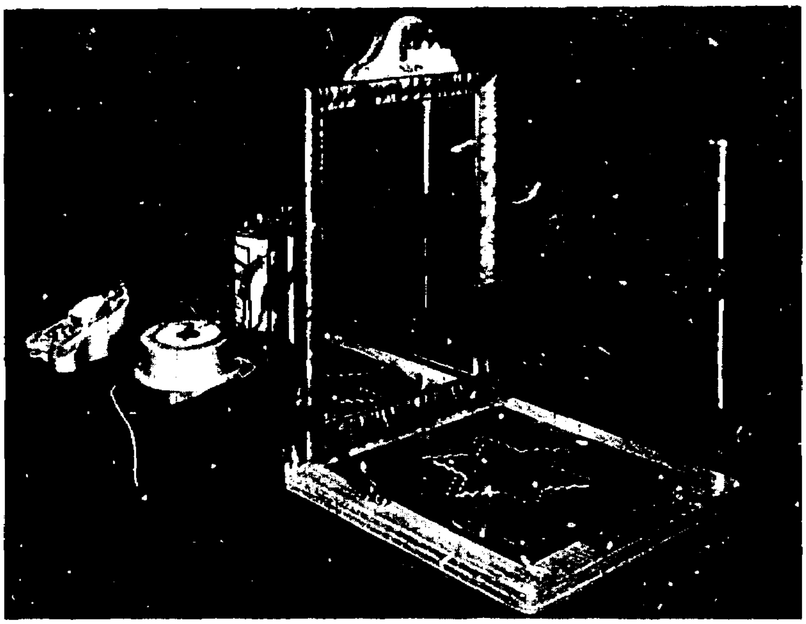


FIGURE XIIc. Photographic picture of mirror-drawing instrument.

Organization Diagram. The relationships of authority and responsibility in political or departmental organization, or the interrelationships of philosophical concepts can frequently be clarified by a diagrammatic presentation. The source of authority and the subordinate officers and branches should be blocked off and arranged in their proper hierarchies. Lines of authority should be solid and lines of co-operation should be dotted. Whenever lines cross unrelated lines of authority, one of the lines should have a "U-shaped" loop at that junction to indicate the unrelatedness.

The accompanying figure is exemplary of a philosophical presentation in diagrammatic form which gives an orderly explanation in more simple form than would ordinarily be possible in paragraph development.

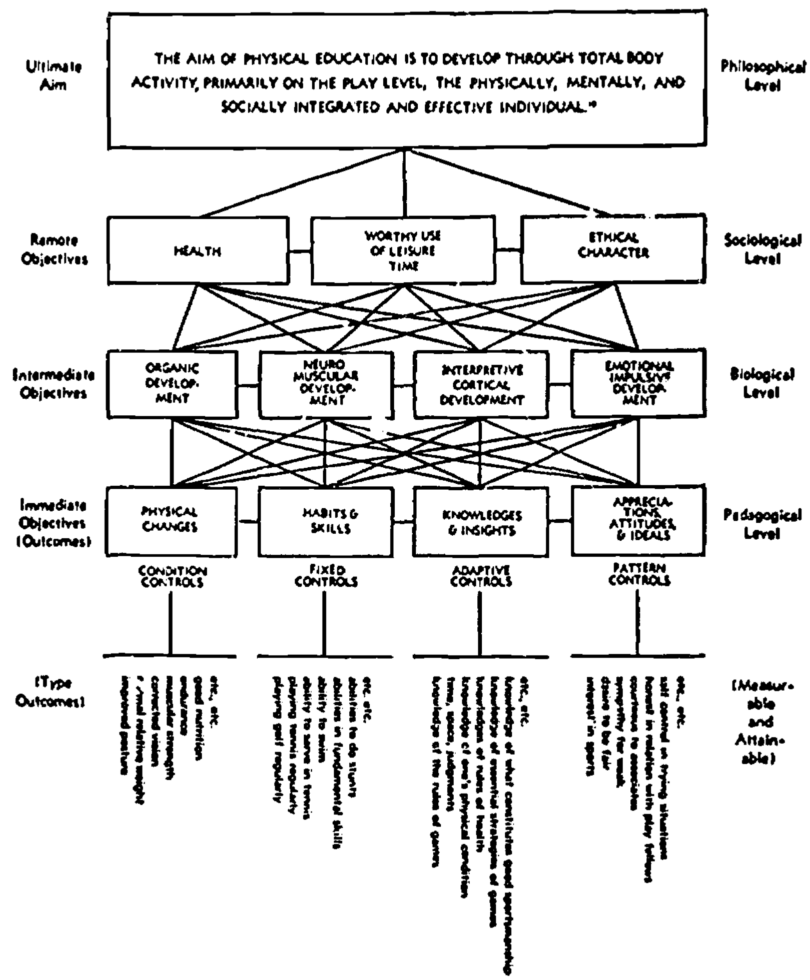


FIGURE XIII. Organization diagram. (Source: *The Physical Educator* 4:4:140; June 1944.)

Maps. As a rule maps of the United States and of the several states can be purchased economically in various stages of detail. The amount of detail needed depends upon the nature of the problem. It is unusual to find district and regional maps readily available.

They must be cut out of larger maps and be reproduced, or larger maps may be used and only the district or region under concern need be utilized. County and city maps are usually obtainable at the court house or city hall. In large cities, cartographers have a wide variety of maps on hand for all purposes.

The cross hatching and shaded areas shown in Figure XIV are of course obtainable in glazine sheets and may be cut to shape and pasted on a larger map which is then reproduced to size. If regions or districts are to be indicated, such as the Eastern District or the Midwest District of the American Association for Health, Physical Education, and Recreation, these demarcations may be made by narrow strips of white paper also cut and applied to the map. As would be expected, the legend shows the higher, intermediate, and lower ranking states in the characteristic studied.

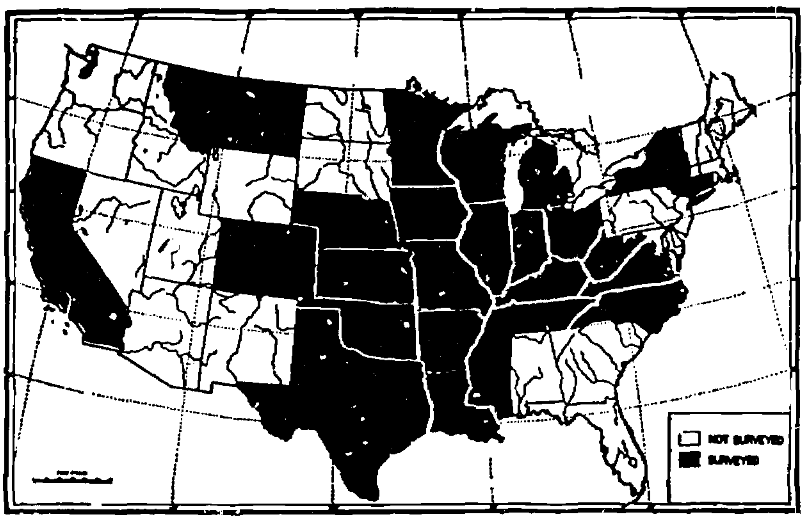


FIGURE XIV. Map (Source: K. W. Bookwalter, *National Survey of Health and Physical Education Programs for Boys in High Schools, 1950-54.* Indiana University, 1953, mimeographed.)

Reproduction of Graphic Material. Abstracts of research reports are a frequent requirement. When more than five or six but less than a hundred copies of material are needed, the original copy and drawings may be typed and traced on ditto or hectograph carbon and the resulting stencil should be sufficiently sharp for this small number of copies.

If several hundred to a thousand copies are needed but once, a mimeograph stencil can be typed or traced as in the original and this number of copies will usually be quite sharp in detail. If exact placing on the page is important, the stencil should be run off on an electric machine rather than a hand-run model.

Should frequent recourse to large numbers of new issue of the material be likely, multilithing, planographing, or other photo-offset processes give values practically as satisfactory as printing and are much cheaper, especially when reproduction of figures is involved. For this type of reproduction, the copy must be a sharp black and white and have no intermediate gray values. Typing with a new black ribbon, backing the page with a new carbon (face up), and tracing all drawings in India ink will assure good results in offset processing. When figures are drawn on co-ordinate paper with blue, green, or sepia lines, the extraneous lines may be filtered out if panchromatic films are used and a filter is attached to the camera. Care must be taken that the size and amount of typing is such that the reproduction can be read. This is especially a problem when drawings are reduced.

When only five to ten copies of figures are needed, sharp detail may be had by photography with Kodalith or similar film. A reasonable substitute process is photostating, though contrasts are not as sharp in this process as in Kodalithing.

FORM IN THE REPORT

Before starting to write up the report, the author should check carefully with the policies and standards required by the editor, publisher, or institution for whom he is writing. By checking on the correct forms for tables and graphs, he will save much time. Suggested order or arrangement, capitalization, and spacing should be decided upon before the report is typed. It is well to make a list of the forms to be followed, so that consistency in form will be assured.

If the report is to be bound in typed or mimeographed form, the left-hand margin should be one and one-half inches. Usually, when the left-hand margin is one and one-half inches, the top margin is also one and one-half inches and the right and bottom margins are one inch.

Footnotes. The most common purposes for which footnotes are used are to cite the authority quoted, to explain a statement in the text, to refer to other sources in the same paper, and to indicate other authorities.

The form for the footnote reference, when referring to the literature, is similar to the form used in the bibliography, except that the exact page reference is given and the indention may differ. If there are many footnotes to be written, the author may wish to use a numbered bibliography and use the numbered footnote in the text as: 1:534; 12:233. In this form, the author indicates references to sources by numbers placed in parentheses at the point of reference. These numbers indicate the proper item in the alphabetically arranged and numbered bibliography and replace the footnote at the bottom of the page. For example, "Deaver's study (19:10)" means the material cited will be found on page 10 in the bibliographical item number 19.

The editorial policies of the *Research Quarterly* (24) require that if regular footnotes are used, the author should identify them in the text with superior figures numbered consecutively throughout the report. Footnotes are to be separated from the text by a horizontal line above and below or typed on a separate page.

For unprinted reports, the footnote is placed at the bottom of the page and separated from the text by a horizontal line. In a lengthy report in which the chapter type of arrangement is used, it is customary to number the footnotes consecutively within the chapter.

When references to the same work appear consecutively on the same page *Ibid* may be used, followed by the page number on which the cited material appears (Example: *Ibid*, p. 20). When a reference has been given in full on a previous page (usually not more than four or five pages back) and intervening references have occurred, the abbreviation *op. cit.* may be used, preceded by the page number or numbers (Example: Deaver, *op. cit.*, p. 12). If more than one reference by the same author is quoted, the name of the book or article must be given. When exactly the same reference is used twice in succession, *Loc. cit.* may be used (Example: Deaver, *loc. cit.*). However, there is a recent trend to use *Ibid* in place of *Loc. cit.* There should be no footnote carry-over from one page to the following page.

Paragraph Development. Sentences within a paragraph should lead from one to the other. It is preferable to avoid constant use of short sentences or of long sentences. Change in style is desirable.

Paragraphs and footnotes are usually indented five or seven spaces. The author should be consistent in the use of the selected indention.

Some hints for good writing form are:

1. The carry-over of a paragraph should have at least two lines to the following page if there has to be a carry-over.
2. Try to eliminate the division of words, but if necessary divide the word according to pronunciation. Do not divide a syllable with a silent vowel. The last word on the page should not be divided.
3. Be consistent in the use of combined words. Some are always hyphenated, while others may or may not be hyphenated.
4. Be consistent in the spelling of words when there is more than one acceptable way to spell a word.
5. The period and comma should be placed inside quotation marks. The period should also be placed inside parentheses or brackets when the matter enclosed is an independent sentence forming no part of the preceding sentence.
6. Seven consecutive words or more of quoted material should be placed within quotation marks or within a quotation paragraph. A quotation of more than three lines should be paragraphed. All quoted material should be quoted *accurately*.
7. It is usually recommended that the underscoring follow the length of the word, with a break before underlining the successive word.

Reference Bibliography. The references in the bibliography should not be numbered, unless the numbers are used for footnote purposes. The references should be alphabetically arranged by author, especially if the references are not numbered. The first line begins at the margin and the carry-over is indented as for paragraph indentions (five or seven spaces). The indention should be the same as for the carry-over in the reference data.

Spacing. Except in certain instances, the usual report should be double spaced except for footnotes, paragraph quotations, bibliography, and special spacing between headings, and for tables and figures. Triple spacing is recommended for spacing between the close of a paragraph and the heading that follows and between the chapter title and the first line of the content. Double spacing is recommended between chapter numbers and chapter titles.

Spacing aids the author in obtaining unity and emphasis of material. Care should be taken to see that the reader obtains the impression desired by the author.

PREPARATION OF MANUSCRIPT FOR PUBLISHER

Following approved procedures in the form of the manuscript will save the author time, energy, and money. All first-rate publishing houses have their own rules of style with which the author should acquaint himself. In general, in preparing the manuscript, type on only one side of the paper, use good quality white paper of 13 to 20 pound weight, and use a page of uniform size, preferably 8½ by 11 inches.

The mechanical features of the report should be consistent throughout, and great care should be exercised to ensure that the report goes to the publisher in the precise form that the author wishes it to be printed. The printer will follow the copy exactly, since he is responsible and must accept the expense for any variations that appear in the proof. For this reason, the author cannot expect the printer to change on his own initiative even those errors which appear to be perfectly obvious.

If the writer wishes to indicate his preference in the use of large or small capitals, italics or bold face type, the following technical marks (19: 56) may be used.

> Underscore the letters or words concerned with (a) three lines (≡≡≡) for capital letters, (b) two lines (===) for small capital letters, (c) one line (———) for italic type, and (d) a wavy line (〰〰〰) for bold faced or black type.

The publisher will return the report to the author for proofreading. It is usually in galley form, a sheet which is much longer than the actual page size that will be used. Since proofreading is a difficult job, it is recommended that the author secure the assistance of another person to read the report aloud while he corrects the galley proof.

Standard proofreaders' marks (13:20, 14:132, 33:184) should be used in making the corrections, and all corrections should be made in the margin. If there is more than one correction in a line, corrections should be placed in order.

It is very important that the author refrain from making changes in the galley proof, unless it is for the correction of an error. Alterations in the galley proof may require resetting of several lines. If it is essential that changes be made, the new material inserted should occupy, if possible, the same amount of space as the original material in the copy.

All corrections should be made immediately, and the galley proof should be returned promptly to the publisher in order to avoid delays in publication.

If it is necessary to estimate a manuscript for printing space, the following procedure may be used. Find the total number of lines and the average length of line in inches on a typical page. Divide the total number of inches by the number of characters on the printed page. The kind of type, pica or elite, should be taken into consideration. Pica type will occupy 10 letters to the inch whereas elite will occupy 12 letters to the inch.

In mailing the manuscript or galley proof, use a tough container that will not tear and that is large enough so that the manuscript is kept flat. It should be labeled carefully, include a return address, and be insured. The author should retain a complete copy, in case of loss in the mail.

BIBLIOGRAPHY

1. Anderson, J. E., and Valentine, W. L. "The Preparation of Articles for Publication in the Journals of the American Psychological Association." *Psychological Bulletin* 41:345-76; 1944.
2. Arkin, Herbert, and Colton, Raymond R. *Graphs, How To Make and Use Them*. New York: Harper & Brothers, 1940.
3. Baker, R. P., and Howell, A. C. *The Preparation of Reports*. New York: Ronald Press Co., 1938.
4. Campbell, William Giles. *A Form Book for Thesis Writing*. Boston: Houghton Mifflin Co., 1939.
5. Cole, Arthur H., and Bigelow, Karl W. *A Manual for Thesis Writing for Graduates and Undergraduates*. New York: John Wiley & Sons, 1934.
6. Cooper, Charles W., and Robbins, Edmund J. *The Term Paper, A Manual and Model*. Stanford Calif.: Stanford University Press, 1934.
7. Crawford, Claude C. *The Techniques of Research in Education*. Los Angeles: University of Southern California, 1928.
8. Davis, W. M. "Preparation of Scientific Articles." *Science* 42:131-34; 1930.
9. Dugdale, Kathleen. *A Manual of Style for Theses and Term Reports*. Bloomington: Indiana University Bookstore, 1955.
10. Ells, Hastings. *Writing a Thesis*. Yellow Springs, Ohio: Antioch Press, 1939.
11. Good, Carter V. "Effective Reporting of Research." *Phi Delta Kappan* 24:178-84; 1941.

12. GOOD, CARTER V., and SCATES, DOUGLAS E. *The Methods of Research*. New York: D. Appleton-Century-Crofts, 1954. (See also GOOD, C. V. *Introduction to Educational Research*. Appleton-Century-Crofts, 1959.)
13. GOOD, WARREN R. *How To Prepare a Term Report*. Ann Arbor: Alumni Press, University of Michigan, 1932.
14. HINKLE, GEORGE, and JOHNSON, FRANCIS R. *The Form for the Term or Research Paper*. Stanford, Calif.: Stanford University Press, 1937.
16. JONES, W. PAUL. *Writing Scientific Papers and Reports*. Dubuque, Iowa: William C. Brown Co., 1946.
17. MELLON, M. G. "Publications: Pitfalls and Problems." *Journal of Chemical Education* 20:534-37; 1943.
18. MEYER, MARGARET, and PELLA, GERMAINE. "The Effect of Hard Laboratory Exercise on the Total and Differential Leucocyte Count of Young Women." *Research Quarterly* 18:271-78; December 1947.
19. MONROE, M. G., and JOHNSTON, NELL B. "Reporting Educational Research." *University of Illinois Bulletin* 22; No. 38, Urbana; University of Illinois, 1925.
20. MUDGETT, BRUCE. *Statistical Tables and Graphs*. Boston: Houghton Mifflin Co., 1930.
21. NELSON, JOSEPH R. *Writing the Technical Report*. New York: McGraw-Hill Book Co., 1947.
22. NEWSOME, N. WILLIAM, and WALK, GEORGE E. *Standards for Thesis Writing*. Scranton, Pa.: International Textbook Co., 1936.
23. NEW YORK UNIVERSITY, SCHOOL OF EDUCATION. *Standards for Written Work*. Scranton, Pa.: International Textbook Co., 1936.
24. REEDER, WARD G. *How To Write a Thesis*. Bloomington, Ill.: Public School Publishing Co., 1930.
25. RESEARCH QUARTERLY. "Guide to Authors." *Research Quarterly*. Published annually in the October issue.
26. RHODES, FRED H. *Technical Report Writing*. New York: McGraw-Hill Book Co., 1941.
27. SASSO, JOHN. "Notes on Preparing a Manuscript for Publication." *Journal of Chemical Education* 20:640-41; 1943.
28. SCATES, DOUGLAS E. "Reporting, Summarizing, and Implementing Educational Research." *Review of Educational Research* 12:558-74; 1942.
29. SCHMITZ, R. MORELL. *Preparing the Research Paper, a Handbook*. New York: Farrar and Rhinehart, 1936.
30. SMITH, HENRY L. *Educational Research*. Bloomington, Ind.: Educational Publications, 1959.
31. SPEAR, MARY E. *Charting Statistics*. New York: McGraw-Hill Book Co., 1955.
32. THURSTON, MARJORIE H. *Preparation and Form of Term Papers and Reports*. Minneapolis: Burgess Publishing Co., 1947.
33. UNIVERSITY OF CHICAGO. *A Manual of Style*. Chicago: University of Chicago Press, 1947.
34. WHITNEY, FREDERICK L. *The Elements of Research*. New York: Prentice-Hall, 1950.
35. ZEISEL, HANS. *Say It with Figures*. New York: Harper & Brothers, 1947.

INDEX

Abstracts, 24, 30
Action current recording, 334
Action research, 433-434, 453-463
  leadership, 458-459
  problem areas, 457-458
  procedure, 459-462
  purpose, 454-457
  value, 453-454
Amplifiers, 153, 155, 159, 163, 164
Analysis of variance, 200-203
Analysis of covariance, 203-203
Anatomy, 327-336
Angoff's equation, 335-336
Annotations, 36
Anthropometry, 300-313
  bibliography, 313-316
  landmarks, 301-303
  physical education, related areas, 308-313
  position and condition of subjects, 302-303
  records, accuracy in, 303
  skills in, 303
  somatotyping, 305-306
  subjects, selection of, 304-306
  unit of measurement, choice of, 304
Apparatus, 344
Appendix, 503-510
Arithmetic mean, 174
Aschenbrenner index, 337
Attention, 378-379
Attitude
  relation to education, 115-116, 119, 358-359
  scale, 116-119
  measurement, 116-119, 331
Attributes, 171
Average or mean, 174
Average deviation, 175

Ballistic movements, 329, 333
Baseball research, 344-345
Beta, 199
Bibliographical form, 503-509
Bibliographies, 33, 63-65, 98, 114, 119-131, 136-158, 145-147, 213, 249-251, 263-264, 269, 276, 290-299, 313-316, 324-325, 345-350, 359-360, 370-372, 393-396, 414-423, 435-437, 449-452, 463-464, 480-481, 500-501, 530-531
Biserial r, 195
Blood, hemoglobin, 366
  pressure, 367, 383
  volume, 366-367
Body build, related studies, 308-303
Body mechanics, 315-324
  bibliography, 324-325
  dynamic positions, evaluation, 318
  feet, 323
  posture tests, 316
  problems, 315, 319
Books, tracing of, 23-24
Buoyancy and specific gravity of the human body, 343-344

Calorimetry, 362
  indirect, 363-365
Cardiac output, 367
Case study technique, 106, 264-268
  bibliography, 269
Center of gravity, 133-143, 317, 333-343
Central tendency, 173-174
Centrifugal force, 344
Chapter headings, 506
Characteristics of a research worker, 5-6, 17-19, 33-48
Checklists, 106
Chi-square, 178, 189-193, 201-203, 206, 209
Chronograph, 153, 156, 160, 161
Chronoscope, 151, 157, 158, 153, 161
Coefficient of alienation, 199
Coefficient of contingency, 195-196
Coefficient of correlation, 193
Coefficient of partial correlation, 196
Confidence
  interval, 73-79, 85-86, 36, 184
  level, 73, 183-184
Computation aids, 310-311
Conclusions, 57-59, 104, 109, 183, 185-187, 203-393, 338, 360, 374, 507-503
Continuous data, 171
Copyright, 35
Correlation
  aids for computing, 310-311
  bi-serial, 195
  contingency, 195-196, 309
  curvilinear (eta), 193
  interpretation, 193
  multiple, 197-198
  partial, 196-197
  Phi, 195
  product-moment, 193
  rank differences (rho), 194, 196
  reliability of, 193-194, 196, 199-200
  tetrachoric, 195
  zero-order, 196
Correlation, errors, 193-194, 195, 198-200
Critical incident, 110-111
Critical ratio, 185-186, 199

Curriculum, 51, 310
  evaluation of, 409-412
  needed research, 413-414
  research in physical education program, 398-414
  research in recreation progr..., 423-435
  research in health education, 437-446

Data, 170-172
  continuous, 171
  discrete, 171
Davis index, 333
Descriptive analysis, 360-362
Dichotomous data, 195
Dictionaries, 24
Difficulty rating, 331
Differences, analysis by, 187-188
Diving, mechanics of, 344
Doctoral theses, distribution, 33
Documentary analysis, 103-109
Dominance, 390-391
Doolittle method, 198, 247-248
Dynamometer, 155, 325

Electrocardiography, 163
Electroencephalograph, 163, 345
Electromyograph, 163, 320, 329, 334, 341, 345
Electrodes, 163, 153, 163, 317
Emotions, 381-385, 408
Empathy, 389-390
Encylopedias and guides, 24-27
Equated groups 280
Equipment, 40, 105, 133-145, 148-160, 336, 364
Ergometer, 164, 361, 363, 355, 367
E.cograph, 156
Error, causes, 72
  intergroup 287-288
  intersubject, 286-287
  of mean, 174, 183-187
  of sampling, 5, 183-184, 286-288
  sources, 288
Eta, 193
Experimental design, 280
Experimental group methods of research, 277-303
  bibliography, 303-306
  patterns for, 290-293
  sampling procedure, 278
Experimental methods, errors, 283-284, 286-288
Experimental methods in laboratory, 368-370

F, 188-189, 201, 206
Facilities, survey of, 354, 357
Factor analysis as a research technique 7, 48, 207-209
  fundamental concepts of, 207
  illustration of, 208
  limitations of, 208
Factorial symbol, 79
Factorial design, 204
Facts and theories, 11-13
Feet, 333-334
Fisher's t value, 188
Flanagan index, 336-337
Footnotes, form for, 509, 527-528
Frequency distribution, 173-174

Gas analysis 163, 364-381
Genetic method, 369-374
Golf research, 345

Graphs, 512-525
  bar, 517-519
  circle or pie cuts, 519-520
  coordinate axes, 512-513
  diagrams, 521
  divided bar, 518-519
  frequency polygon, 513-514
  histogram, 514-515
  line, 515-517
  maps, 524-525
  multiple line, 516
  organization diagram, 523-524
  pictograph, 520-521
  pictures, 521-523
  sketch picture, 522
Gymnastics, mechanics of, 133, 338-339, 344

Health education curriculum, 437-446
  bibliography, 446-452
  historical trends, 439-441
  needed research, 445-446
Health factors in posture, 320-321
Health knowledge tests, 319-320, 443
Hearing, 385
Heart and circulation, 363-367
Heart rate, 367
Historical research methods, 459-480
  bibliography, 480-481
  criticism, 473-477
  interpretation, 478-480
  purpose and scope, 465-469
  sources, 469-475
Historical trends
  in physical education curricular research, 398-401
  in recreation program, 425-428
  in health education, 439-441
Horizontal analyses, 54-65, 64-70
Hotelling method, factor analysis, 207
Hypothesis, 5, 184-188, 191-193, 202, 358, 401-405
  null, 187, 183, 114, 185, 198, 381-384
H-score, 178

Illustrations 30, 142-144, 510-520
Indentations in typing, 533
Index of discrimination, 326-330
Indexes, 23-27, 33-34
Inductive process, 483-484
Instruments
  in testing, 345, 506, 303
  in surveying, 353
Interpretation of data, 171-173, 358-360
Interview technique, 100-103

Jumping, mechanics of, 331-332

Kinesiology, experimental, 326-345
  ballistic movements, 328, 339
  basic requirements for, 327-331
  bibliography, 345-350
  fundamental activities, 336-345
  methods, 331-336
  movement recording systems, 332-333
  muscle action recording systems, 333-335
  time analysis, 333
Kinesthesis, 48, 160, 385-388
Knowledge tests
  construction, 315-335
  criteria, 330-331
  difficulty rating, 331
  form of questions, 319-323

non-functioning foils, 231-232
norms, 236-237
reliability, 234-236
revising, 233-234
use, 112-113, 220-231
validity, 224-231
Kuder-Richardson formula, 234
Kurtosis, 178, 193

Laboratory
administration, 166
costs, 148-149
curricular relations, 148-149, 353
equipment, 148-169, 300, 336
facilities, 166-169
Law, 5, 11, 15
Laws of chance, 78, 180-181, 183
Law of parsimony, 281
Law of probability, 77, 176-178
Learning, 350-359, 374-377, 405-408
aids, 355-356, 413-413
bibliography, 359-360
curves, 375-377
Library classification systems, 24-25
inter-library loan, 25
Library research methods
abstracts, 24
bibliography, 38
dictionaries and glossaries, 24
encyclopedias and guides, 24
indexes, 23-27, 33-34
purpose, 20-21, 53
suggestive headings for, 32-34
List of figures, 505
List of tables, 505
Literature, research, how to locate, 34 35
Logical analysis, 434, 437-463, 431-495
Lung ventilation
rate and amplitude, 364-366, 383
volume, 163-164, 364-366

Manuscript, preparation for publisher, 529-530
Margins, 526
Matched samples, 188, 288, 370
Mean or average, 174
error of, 185, 183-187
Median, 172-174
Meredith growth records, 310, 443
Metabolism, 163-164
Microcards, 33, 509
Microfilm, 26, 36
Mode, 173
Motivation, 369, 379-381
Movement
range, 316, 337, 335
recording systems, 333-335
speed of, 130-133, 150-152, 333
Multiple correlation, 247-249
Muscular tension, 333, 383-383
Muscle action recording systems, 333-335

Nature and scope of research, 2-3, 6 9, 11-16, 40-43
Non-parametric methods, 189, 309-310
Normal curve, 178-179
Normal distribution, 73, 181
Normative method, 7, 370
Norms, meaning, 236-337, 321
Number of cases, 81-83, 87, 88, 93, 379

Objectivity, 104
Observation technique, 103-106
Oscillograph, 153

Partial correlation, 196-197
Pearson product moment correlation, 193
Percentiles, 179
Periodicals, 30-33
Personality, 391-393, 405
Phi coefficient, 239
Philosophical research method
applied to curriculum, 408-409
bibliography, 500-501
illustrations of, 484-489, 496-500
limitations of, 484-488 494
procedure, 489-495
Photographical and cinematographical research
analysis, 134-144, 338-340, 341, 344
bibliography, 145-147
filming, 145, 337
illustrations, 143-144
spatial relationships, 129-130
timing, 130-133
typical data, 135-143
use in physical education, 138-139, 133-135, 337, 331-333
Photography, types of, 128, 134, 331, 336
Physical capacity tests, 164, 362-365
Physical education curriculum
bibliography, 414-433
historical trends, 399-401
needed research, 413-414
research related to, 401-413
Physical growth, trends of
Meredith charts, 310
Wetzel grid, 44, 309, 443
Physical performance tests, 164-166, 305, 313, 363-365
Physical status, related studies, 373, 401 403
Physiological factors in measurement errors, 105, 337, 317, 369
Physiological laboratory research, 360-370
bibliography, 370-372
blood, hemoglobin, 366
blood, pressure, 367
blood, volume, 356-368
cardiac output, 367
electrocardiography, 163
gas analysis, 364-365
heart and circulation, 356-368
heart rate, 367
lung ventilation, rate and amplitude, 364-366
lung volume, 163-164, 364-366
mechanical work, 361-363
physiological work, 363-365
respiratory and metabolism energy exchange, 365-366
work, 361-365
Physiological research, applied, 336-330, 360-361
Play, 391
Plagiarism, 508-509
Population, 71-93
Posture
bibliography, 334-335
health factors, 330-333, 324
objective measurement of, 315-317
related problems, 313-324
sampling, 317-313
structural factors, 317
validation, problems of, 316, 318 330, 333 336
Prediction, 9-11, 193-200, 373
Preface, 505
Probable error, 8, 176
Probability theory, 77-78, 177

Problems, selecting and planning, 76, 253-254, 266, 272-273, 278-279, 370, 459-461, 469, 490
Profession of physical education, 1-3, 51
Projective tests, 118-119, 357, 392
Psychological factors in measurement errors, 105, 369
Psychological laboratory research, 373, 378-382
  attention, 378-379
  bibliography, 393-396
  dominance, 390-391
  emotion, 381-382
  empathy, 389-390
  hearing, 385
  kinesthesis, 385-388
  learning, 374-377
  motivation, 379-381
  muscular tension, 382-383
  personality, 391-393
  play, 391
  projective tests, 118-119, 357, 393
  reaction time, 157-159
  rhythm, 388-389
  sensations, 383-385
  speed of movement, 130-153, 150-153, 383
  transfer, 377-378
  vision, 383-385
Psychological research applied, 373-373

Quartile, 175
Questionnaire technique, 100-103
Quotations, 36

R, 197-198
r, 193-194
Random error, 77-79, 84, 193, 381, 386, 383
Range, 175
Ratings, 108-108, 339-340
Reaction time, measuring, 157-159
Recording systems, 333-335
  muscle action, 333-335
  pneumatic, 333-334
  thread and rubber band, 345
Recreation, 423-435
  co-operative research, 427-428
  historical trends, 425-427
  problems and limitations, 428-430
  research clinics, 433-434
  needed research, 434-435
Reference bibliography, 538
Regression equations, 198-199, 307, 343
Relationship analysis, 191-198, 358, 364, 247-368, 370-371, 491
Reliability, 106-107, 111-112, 176, 183, 193, 136
Reliability coefficient, 254-236, 344
Reliability, index of, 234-236
Repeated measurement, 334, 371, 280, 285, 303, 310, 316
Research
  action, 4
  administrative, 355-356
  anthropometric, 300-313
  applied, 2-4
  basic, 2-4
  experimental, 8-9, 377-398
  functions of, 43-49
  future emphasis, 16-17
  general procedures, 53-62
  historical, 6, 14, 469-480
  meaning of, 454-457
  methods, scientific, 6
  nature and scope of, 2-3, 6-9, 11-16, 40-42
  observational or clinical, 6-7, 103-106, 203
  philosophical, 482-500
  physiological, applied, 360-370
  problematical areas, 16-17, 305-306, 315-324, 351-358, 370, 385, 390, 391, 393, 413-414, 434-435, 445-446, 457-458
  psychological, applied, 372-393
  pure, 3-4
  scientific method in, 3, 9-16, 170-171, 466-469, 482-483
  steps in, 6, 54, 64-70
  unique nature of, in physical education, 2-4, 150
  why, 4-6, 17-19, 454, 483
Research report
  bibliographical and footnote form, 21, 527-529
  characteristics of, 502-503
  function of, 260-363
  major divisions of, 504-510
  mechanical features of, 503-504, 526-530
  preliminaries to, 503-504
  proofing, 529-530
Research literature, how to locate, 33-35
Research methods in the mechanics of sports and physical education activities, 135-143, 336-345
  bibliography, 145-147, 345-350
Research problems
  areas according to educational emphasis, 43, 49-51
  horizontal analysis of, 56-62, 64-70
  plan of, 54
  rating of, 48-49
  selection of, 49-54
Research worker
  characteristics of, 18-19, 40-41, 44-48
  errors of, 13, 37, 39
Respiration, 343-344
Rho, 194, 309
Rhythm, 388-389
Running, mechanics of, 338-340

Sampling, errors in, 73, 78, 81-83, 183, 185, 188-189, 309-210, 238, 386-387
Sampling, 71-98, 180-183, 358
  cluster, 93-95
  multi-stage, 96
  non-probability, 97-98
  random, 77-79, 83-87, 180-183, 384, 293-393
  ratio, 90
  replicate, 96-97
  representative, 9
  sequential, 96
  size of, 41, 37, 88-89, 93, 188-189, 279
  stratified, 37-93
  systematic, 95-96
Scale scores, 179-180
Science, purpose and method of, 9-10, 170-171
Score cards, 58-59, 109-110
Scores, interpretations of
  percentile, 179
  raw, 171-172, 258-359
  T-score, 180
  Z-score, 179
Selecting and planning problems, 39-43
  bibliography, 62-93
Sensations, 383-388
Sigma, 175

Significance levels, 282-286
  errors, 282-289
Skewness, 174, 178, 193
Sociometrics, 131-236, 353
  interpretation, 136
  sociogram, 135
  tests, 133-136
Somatotyping, 305-308
Sources, 36, 469-475
Spacing, 505, 510, 513, 528-529
Spearman-Brown formula, 196, 235
Spearman's rank order correlation, 194, 209
Specific gravity, 138-143, 331
Speed of movement, 130-133, 150-152, 332
Sports and physical education activities
  cycling, mechanics of, 336
  jumping, 331-333, 335
  locomotor activities, 329, 338-342
  running, mechanics of, 337-340
  starting, mechanics of, 336
  throwing, mechanics of, 333, 335, 344-345
  swimming, mechanics of, 144, 342-344
Standard deviation, 170-176
Standard error, or mean, 78, 91-93, 94-95
Standard scores, 227
Starting mechanics, 336-336
Statistical methods
  aids in computation, 210-211
  central tendency, 173-174
  coefficient of contingency, 195-196, 209
  comparative analysis, 171
  descriptive analysis, 171
  errors, 173
  graphic methods, 513-535
  group scores, comparison of, 280-284
  interpretation of data, 170-173, 174
  normal curve, 176-179, 181
  partial correlation, 196-197
  percentiles, 179
  product-moment correlation, 193
  rank order (rho), 194, 209
  sampling, 180-183
  scale scores, 179-180, 227
  small sample technique, 188-189
  types of data, 171-173
  variable, 171
  variability, 174-176
Step interval, 179
Stick figures, 135-136
Strain gauge, 165, 139
Structural factors, 305-308, 310, 323-324
Survey methods
  analytical, ?
  aspects of professional areas, 252-253
  bibliography, 263-264
  criteria for, 252-253
  interpretation, 256-263
  scope of, 254-256
  steps in, 253-263
Swimming, mechanics of, 144-145, 342-344, 363
Swineford index, 236
Systematic error, 266

t, 85, 188, 202
Table of contents, typing of, 505
Tables, examples of, 510-513
Tables, typing of 510-513
Tabular analysis, 507
Tabular and graphic presentation, 510-526
Tachistoscope, 165-166
Tennis research, 345
Tensiometer, 155, 325
Tension, muscular, 382-383
Test construction and analysis methods
  bibliography, 249-251
  difficulty rating, 221
  knowledge tests, 112-113, 215-227
  physical performance tests, 111-113
  physical performance, criteria, 238-241
  physical performance tests, steps in, 227-249
  reliability, 224-226, 242-245
  validity, 111-112, 228-231, 245
Tetrachoric correlation, 195
Theory, 11, 72-79, 180-183
Thurstone method, factor analysis, 207-209
Time
  reaction time, 157-159
  analysis of, 333-333
  recording of, 130-133, 150-153, 157-159, 165, 333
Title page, typing of, 504-505
Track, studies in, 326-328, 365
Transfer, 377-376
Treadmill, 164, 339-340, 331, 362, 333
True score, 176
T-score, 179, 227
Tumbling, mechanics of, 344
Typing aids, 222-530

Underscoring in typing, 529

V, 176
Validity, 111-112, 228-231
Variable, 171
Variability, 174-179
Variance in population, 84, 89-91
Vision, 333-335
Vetaw index, 329

Walking, mechanics of, 340-342
Weights, mechanics of, 349
Wetzel nutritional grid, 44, 308, 309, 443
Wherry-Doolittle, 198, 247-248
Work, 236, 261-235
Writing the research report, 503-530
  bibliography, 330-531
Written tests, 112-113, 215-227

z-scores, 179
zero-order, 196